

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2004, 21:15:30 ; Search time 112 Seconds
(without alignments)
9959.387 Million cell updates/sec

Title: US-10-063-732-119
Perfect score: 2010
Sequence: 1 ggaaaaactgtctctctctg.....aattgtacattttctaatt 2010

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	493.2	24.5	615	4	US-09-621-976-1835 Sequence 1835, App
2	175.8	8.7	708	4	US-09-620-312D-986 Sequence 986, App
3	159.2	7.9	1705	4	US-09-205-258-106 Sequence 106, App
4	158.6	7.9	1665	4	US-09-300-958A-22 Sequence 22, Appl
5	150	7.5	1253	4	US-09-673-355A-71 Sequence 71, Appl
6	149	7.4	1722	4	US-09-489-847-93 Sequence 93, Appl
7	134.8	6.7	3483	4	US-09-886-683A-3 Sequence 3, Appl
8	134.8	6.7	3483	4	US-09-130-491-3 Sequence 125, App
9	129	6.4	1380	4	US-09-489-847-125 Sequence 45, Appl
10	124.8	6.2	1524	4	US-09-663-600A-45 Sequence 139, Appl
11	124	6.2	1400	4	US-09-663-600A-139 Sequence 55, Appl
12	94.8	4.7	413	3	US-09-188-930-55 Sequence 55, Appl
13	94.8	4.7	413	4	US-09-312-283C-55 Sequence 55, Appl
14	89.8	4.5	441	4	US-09-404-879A-104 Sequence 104, App
15	89.8	4.5	441	4	US-09-338-933-104 Sequence 104, App
16	89.8	4.5	441	4	US-09-215-681-104 Sequence 104, App
17	89.8	4.5	441	4	US-09-216-003A-104 Sequence 104, App
18	88.4	4.4	363	2	US-08-966-316-8 Sequence 8, Appl
19	78.6	3.9	1353	4	US-09-724-864-32 Sequence 32, Appl
20	64.8	3.2	413	4	US-09-621-976-360 Sequence 360, App
21	58.6	2.9	1165	4	US-09-603-552-2 Sequence 2, Appl
22	52.8	2.6	615	3	US-08-998-416-186 Sequence 186, App
23	52.2	2.6	731	1	US-08-451-405A-2 Sequence 2, Appl
24	52.2	2.6	640681	4	US-09-790-988-1 Sequence 1, Appl
25	52	2.6	636	3	US-08-998-416-1137 Sequence 1137, App
26	52	2.6	19124	2	US-08-487-826B-13 Sequence 13, Appl
27	50.8	2.5	6113	4	US-10-204-708-14 Sequence 14, Appl

28	50.2	2.5	10467	4	US-10-204-708-2	Sequence 2, Appl
29	49.6	2.5	10640	4	US-09-417-485D-5	Sequence 5, Appl
30	49.4	2.5	15062	4	US-09-004-838-89	Sequence 89, Appl
31	49.4	2.5	20674	4	US-09-641-638-651	Sequence 651, App
32	48.2	2.4	1511	1	US-07-991-867B-8	Sequence 8, Appl
33	48.2	2.4	1511	1	US-08-107-755A-8	Sequence 8, Appl
34	48.2	2.4	1511	2	US-08-544-332-8	Sequence 8, Appl
35	48.2	2.4	1511	4	US-09-370-861A-8	Sequence 8, Appl
36	47.8	2.4	832	4	US-09-621-976-2813	Sequence 2813, Ap
37	47.6	2.4	427	4	US-09-603-552-8	Sequence 8, Appl
38	46.8	2.3	3739	4	US-09-409-604-3	Sequence 3, Appl
39	46.8	2.3	640681	4	US-09-790-988-1	Sequence 1, Appl
40	46.6	2.3	6040	4	US-10-204-708-70	Sequence 70, Appl
41	46.6	2.3	6124	4	US-08-213-419B-3	Sequence 3, Appl
42	46.2	2.3	5852	1	US-07-867-106-2	Sequence 2, Appl
43	46	2.3	660	1	US-07-991-867B-32	Sequence 32, Appl
44	46	2.3	660	1	US-08-107-755A-32	Sequence 32, Appl
45	46	2.3	660	2	US-08-544-332-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-09-621-976-1835
Sequence 1835, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621, 976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 1835
LENGTH: 615
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 173..511
NAME/KEY: sig_peptide
LOCATION: 173..409
OTHER INFORMATION: Von Heijne matrix
OTHER INFORMATION: score 4.30000019073486
OTHER INFORMATION: seq YDSLALSPDLQA/AR
US-09-621-976-1835

Query Match	24.5%;	Score 493.2;	DB 4;	Length 615;
Best Local Similarity	98.3%;	Pred. No. 7.4e-119;		
Matches 529;	Conservative 1;	Mismatches 4;	Indels 4;	Gaps 3;
QY	1	GGAAAAACTGTTCTTCTTCTGTGGACAGAGAAACCCTGCTTCAAGCAGAGTAGCAGTTC	60	
Db	77	GGA AAAA CTGTTCTTCTTCTGTGGACAGAGAAACCCTGCTTCAAGCAGAGTAGCAGTTC	136	
QY	61	CGAGTCCAGCTGGCTAAACTCATCCAGAGGATATGGCAACCCATGCTTAGAATC	120	
Db	137	CGAGTCCAGCTGGCTAAACTCATCCAGAGGATATGGCAACCCATGCTTAGAATC	196	
QY	121	GCTGGGCTGTTCTTCTTCTGTGGATGTTGGGCACTGGCTGTCACTGTCATGCT	180	
Db	197	GCTGGGCTGTTCTTCTTCTGTGGATGTTGGGCACTGGCTGTCACTGTCATGCT	256	
QY	181	CAGTGAGAGTGTGCGCCTTCATGAAAACAACATCGTGGTTTGAAGACTTCTGGAA	240	
Db	257	CAGTGAGAGTGTGCGCCTTCATGAAAACAACATCGTGGTTTGAAGACTTCTGGAA	316	
QY	241	GGA CTGTGATGAATTCGTGAGGAGGCTAACATCAGATGCAGTGCAAAATCTATGAT	300	

```
Db      317  GGAAGTGGATGATATGCGGTGAGGAGGCTAAATCAGATGAGTGCAGAAATCTATGAT 376
QY      301  TCCCTGCTGCTCTTTCTCCGACCTACAGGAGCCAGAGAGTGTGTGCTGCTCC 360
Db      377  TCCCTGCTGCTCTTTCTCCGACCTACAGGAGCCAGAGAGTGTGTGCTGCTCC 436
QY      361  GTGATGCTCTTCTGCTTCATGATGAGCCATCTTGGCATGAATGACACAGGTGACG 420
Db      437  GTGATGCTCTTCTGCTTCATGATGAGCCATCTTGGCATGAATGACACAGGTGACG 495
QY      421  GGGGACATGAGAGGTGAAGGCTCACATTCCTGAGCGGCTGGAATCATCTTCATCATC 480
Db      496  GGGGACATGAGAGGTGAAGGCTCACATTCCTGAGCGGCTGGAATCATCTTCATCATC 555
QY      481  ACGGCGATGGTGTGCTCATCCCTGTGAGCTGGGTTGCCAATGCCATCATCAGAGATT 538
Db      556  GCGGCGATGGTGTGCTCAT- CCTGTGAGCTGGGTTGC- AATGCATCATCAGAGATT 610
```

RESULT 2

```
US-09-620-312D-986
; Sequence 986, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yundong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pc_fl_genes Version 1.0
; SEQ ID NO 986
; LENGTH: 708
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (46)..(708)
US-09-620-312D-986
```

```
Query Match      8.7%; Score 175.8; DB 4; Length 708;
Best Local Similarity 57.4%; Pred. No. 3.8e-36;
Matches 337; Conservative 0; Mismatches 247; Indels 3; Gaps 1;

QY      66  TCCAGCTGGCTAAACTATCCAGAGATATGGAACCCATGCTTGAATGCTGG 125
Db      15  TCCTTCGAGTCAGCTCCTTCAACCTGCGCATGGCCTCTGCCGAATGAGATCTGGG 74

QY      126 GCTGTTCTTGTGTGTGTAATGTTGGGACAGTGGCTGTCACTGTCACTGCTGAGTG 185
Db      75  AGTCGTCTGACACTGCTGGGCTGGGTGAATGGCTGTCTCTGTGCTGCCCATGTG 134
```

```
QY      186  GAGAGTGGCCCTTCATTGAAAACACATCGTGCTTTTGAACCTTCTGGAGAGACT 245
Db      135  GAAGTGACCGCTTTCATCGCAACAGCATCGTGCTGGCCAGGTGTGTGGAGGGCT 194
QY      246  GTGATGAATTCGTGAGGAGGCTAACATCAGATGCAGTCAAAATCTATGATTCCT 305
Db      195  GTGATGCTCTCGTGTGTGAGAGCACCGGCCAGATGAGTGAAGTGAAGTACACT 254
QY      306  GCTGCTCTTCTTCGAGCTACAGGAGCCAGAGGACTGATGTGCTGCTCCGTAT 365
Db      255  GCTGCGCTGACACAGAGCTGACAGGCTGACAGTCCCTCTGTGATCGCCCTCTTGT 314
QY      366  GTCCTTCCTGCTTTCATGATGAGCCATCTTGCCATGAATGACACAGGTGACGGGGA 425
Db      315  GGCCTGTTGCGCTTGTGCTGCTAACCCTGCTGGGGCCAAAGTGAACACCTGTGTGAGGA 374
QY      426  CAATGAGAGGTGAAGGCTACATTTCTGCTGACGGCTGGAATCATCTTCATCACGGG 485
Db      375  GA--AGGATTCGAAGGCCCGCTGTGCTCACTCTGGATGTCTTGTCACTCAGG 431
QY      486  CATGTGTGCTCATCCCTGTGAGCTGGTGGCAATGCCATCATCAGAGATTCTATAA 545
Db      432  GGTCTGACGCTAATCCCGGTGTGCTGACGGCGCATGCCGTCACTCCGAGACTTCTATA 491
QY      546  CTCATATGATGATTTGCCCAAAACGTGAGCTTGAGAGAGCTCTCTACTTAGATGAC 605
Db      492  CCCCCTGTGGCTGAGGCCCAAAAGCGAGCTGGGGGCTCCCTCTACTTGAGCTGGGC 551
QY      606  CACGCGACTGTGCTGATTTGTTGAGAGAGCTCTGTTCTGCTGCGTTT 652
Db      552  GGCTCAGGCTTTTGTGTGCTGGGTGGGGGCTGTGCTGTGCACTT 598
```

RESULT 3

```
US-09-205-258-106
; Sequence 106, Application US/09205258
; Patent No. 6525174
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; FILE REFERENCE: P2007P1
; CURRENT APPLICATION NUMBER: US/09/205,258
; EARLIER FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/11422
; EARLIER FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,885
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,375
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,881
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,880
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,896
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,020
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,876
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,895
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,884
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,894
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,971
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,882
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,899
; EARLIER FILING DATE: 1997-06-06
```

```

; EARLIER APPLICATION NUMBER: 60/048, 893
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 900
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 901
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 892
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 915
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049, 019
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 972
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 916
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049, 373
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 875
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049, 374
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 917
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 949
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 883
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 897
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 898
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 962
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 963
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 877
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048, 878
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/070, 923
; EARLIER FILING DATE: 1997-12-18
; EARLIER APPLICATION NUMBER: 60/092, 921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/094, 657
; EARLIER FILING DATE: 1998-07-30
; NUMBER OF SEQ ID NOS: 1227
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 106
; LENGTH: 1705
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (724)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-205-258-106

```

```

Query Match          7.9%; Score 159.2; DB 4; Length 1705;
Best Local Similarity 56.6%; Pred. No. 1.3e-31;
Matches 315; Conservative 0; Mismatches 239; Indels 3; Gaps 1;
QY 92 GGATAATGGCAACCCATGCCCTTAGAATCGCTGGGCTGTTCTTGTTGTTGATG 151
DB 204 GAACATGGCCTCCATGAGGCTACAGTAATGGGCATCGCGCTGCCCTGAGCTGGC 263
QY 152 TGGGCAAGTGGCTGTCACTGTCACTGCTCAGTGGAGAGTGTGCGCTTCATGAAACA 211
DB 264 TGGCCGTATGCTGTGTGCTGGCGCTGCCCATGTGGCGCGGTGACGCGCTTCATCGGACGA 323

```

```

QY 212 ACATCGTGTGTTTTGAAACTTCTGGGAGGACTGTGATGAATTGCGTAGGCGACTA 271
DB 324 ACATTGTACCTTCGACAGACCATCTGGGAGGGCCTATGATGAACCTGCGTGCAGAGCA 383
QY 272 ACATCAGATGCAAGTCAAAATCTATGATTCCTCGTGGCTCTTCTCCGACCTACAGG 331
DB 384 CCGGCAGATGCAAGTCAAGTGTACGACTGCTGTGGCACTGCCGACGAGACCTGCAGG 443
QY 332 CAGCCAGAGACTGATGTGTGCTGCTTCCGTGATGTCCTTCTTGACTTCATGATGCCA 391
DB 444 CGGCCCGCGCCTCGTCATCATCAGCATCATCGTGGCTGCTCTGCGGCTGCTGCCG 503
QY 392 TCCTTGCAATGAATGCACCAAGTGCACGGGACATGAGAAGTGAAGGCTCACATTG 451
DB 504 TGCTGGGGGCAAGTGTACCAACTGCTGAGG---ATGAAGCGCCAGGCCAAGACCA 560
QY 452 TGCTGACGGCTGGAATCATCTTCATCATCAGCGGCAATGCTGTCTCATCCCTGTAGCT 511
DB 561 TGATCGTGGCGGCGCTGTGTCTCTGTGGCCGCTTATGTGATAGTCCGCTGTCT 620
QY 512 GGGTGGCAATGCCATCATCAGAGATTCTATTAATCAATAGTGAATGTTGCCCAAAAC 571
DB 621 GGACGGCCCAACATCATCCAGACTTCTACAATCCGCTGTGTGCGCTCCGGGCAAGAC 680
QY 572 GTGACCTTGAGAGAGCTCTCTACTTAGATGAGACCAAGGCACTGCTGTGATGTTGGAG 631
DB 681 GGGAGATGGGTGCTCTGCTCTTACGTGGCTGGCGCCGCTCCGNGCTGTGCTCTTGCGG 740
QY 632 GAGCTGTGTTGCTGCTGC 648
DB 741 GGGGCTGCTTTGCTGTC 757

```

```

RESULT 4
US-09-300-958A-22
; Sequence 22, Application US/09300958A
; Patent No. 6495319
; GENERAL INFORMATION:
; APPLICANT: McClelland, Michael
; APPLICANT: Welsh, John
; TITLE OF INVENTION: Reduced Complexity Nucleic Acid Targets and Methods of
; FILE REFERENCE: P-PH 3457
; CURRENT APPLICATION NUMBER: US/09/300,958A
; CURRENT FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/083,331
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: 60/098,070
; PRIOR FILING DATE: 1998-08-27
; PRIOR APPLICATION NUMBER: 60/118,624
; PRIOR FILING DATE: 1999-02-04
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1665
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-300-958A-22

```

```

Query Match          7.9%; Score 158.6; DB 4; Length 1665;
Best Local Similarity 56.6%; Pred. No. 1.8e-31;
Matches 315; Conservative 0; Mismatches 239; Indels 3; Gaps 1;
QY 92 GGATAATGGCAACCCATGCCCTTAGAATCGCTGGGCTGTTCTTGTTGTTGATG 151
DB 178 GAACATGGCCTCCATGAGGCTACAGTAATGGGCATCGCGCTGCGCTGCTGAGCTGGC 237
QY 152 TGGGCAAGTGGCTGTCACTGTCACTGCTCAGTGGAGAGTGTGCGCTTCATGAAACA 211
DB 238 TGGCCGTATGCTGTGTGCTGGCGCTGCCCATGTGGCGCGGTGACGCGCTTCATCGGACGA 297

```

OY	212	ACATCGTGCTTTTGGAAAACCTTCGGGAAGGACTGTGATGAATTGCCGTAGGCAGGCTA	271
Dd	298	ACATTGTCACTCCGACAGACCATTGGGAGGGCCTATGATGAATCGCGTGTGCAGAGCA	357
OY	272	ACATCAGGATGCAGTGC AAAATCTATGATTCCCTGCTGCTCTTCTCCGACCTACAGG	331
Dd	358	CCGGCCAGATGCAGTGC AAGGTATACGACTGCTGTGGCACTGCCGACGACCTGCAGG	417
OY	332	CAGCCAGAAGACTGATGTGTGCTGCTCCGTATGTCCCTTCCTTGCGCTTCATGATGGCCA	391
Dd	418	CGGCCCGCGCCCTCGTCATCATCAGCATCATCGTGTGCTGTGGCGTGTGCTGTCCG	477
OY	392	TCC TTGGCATGAAATGC ACCAGGTGC ACGGGACAATGAGAAGTGAAGGCTCACATT C	451
Dd	478	TG GTGGGGGGCAAGTGTACCAACTGCTTGAGG -- ATGAAGCGCCAAGGCCAAGACCA	534
OY	452	TGCTGACGGCTGGAATCATCTTCATCATCA CGGCATGGTGTGCTCATCCCTGTGAGCT	511
Dd	535	TGATCGTGGCGGGCGTGTGTTCCTGTGGCCGGCCTATGGTAGTAGTGCCGGTGTCT	594
OY	512	GGGTGGCCATGCCATCATCAGAGATTTCTATACTCAATAGTGAATGTTGCCAAAAAC	571
Dd	595	GGACGGCCCCAACATCATCCAGA ACTTCTACAATCCGCTGGTGGCCTCCGGCAGAGA GC	654
OY	572	G TGAGCTTGAGAAGCTCTCTACTTA GGA TGGACCA CGGCA CTGGTGTGATTTGTTGAG	631
Dd	655	GGGAGATGGGTGCTCTGCTCTACGTGGCTGGGCCCGCTCCGGCTGCTGTCTGTTGGCG	714
OY	632	GAGCTCTGTCTGCTGC	648
Dd	715	GGGGGCTGCTTTGCTGC	731

```

RESULT 5
US-09-673-395A-71
; Sequence 71, Application US/09673395A
; Patent No. 6620923
; GENERAL INFORMATION:
; APPLICANT: SPECHT, THOMAS
; APPLICANT: HINZMANN, BERND
; APPLICANT: SCHMITT, ARMIN
; APPLICANT: PILARSKY, CHRISTIAN
; APPLICANT: DAHL, EDGAR
; APPLICANT: ROSENTHAL, ANDRE
; TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM UTERUS TUMOR TISSUE
; FILE REFERENCE: ALBRE-12
; CURRENT APPLICATION NUMBER: US/09/673,395A
; CURRENT FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 637
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 71
; LENGTH: 1253
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-673-395A-71

```

	Query Match	7.5%;	Score 150;	DB 4;	Length 1253;
	Best Local Similarity	55.1%;	Pred. No. 2.7e-29;		
	Matches 314;	Conservative	0;	Mismatches 255;	Indels 1; Gaps 1;
Oy	115	GAAATCGCTGGGCTGTTCCTTGGTGGTGGAAT-GGTGGGCACAGTGGCTGTCACTGT	173		
Db	363	GCAGTTGCTGGGGCTTCTCCATGGGCCCTGTCTGGGCTGGGTGGGTCTGGTGGCCTGCACC GC	422		
Oy	174	CATGCCCTAGTGGAGAGTGTGGGCTTCATTGAAAACAACATCGTGGTTTTTGA AAACTT	233		
Db	423	CATCCCCGAGTGGCAGATGAGCTCTATGCGGGTGACACAACATCATCACGGCCCA GGCCAT	482		
Oy	234	CTGGGAAGGACTGTGATGTAATTGGGTGAGGCGAGCGCTA ACATCAGGATGCAGTGCA AAAT	293		
Db	483	GTACAAGGGGCTGTGATGACTGGCTCACGCAGAGCACC GGGA TGATGAGCTGCCAA AAT	542		
Oy	294	CTATGATTCCTTGCTGCTCTTTTCTCCGGACCTACAGGACCCAGAGGACTGATGTGTGC	353		

Db	543	GTACGACTGGTGCTCGCCCTGTCGGCCCTGACGCCACTCGAGCCCTAAATGGTGGT	602
QY	354	TGCTTCCGATGATGTCCTTCTTGGCTTCATGATGGCCATCCTTGGCATGAATGCACG	413
Db	603	CTCCCTGGTGCTGGGCTTCCCTGGCCATGTTTGTGGCCACGATGGCATGAAGTGCACGCG	662
QY	414	GTGCACGGGGACATAGAGAAGTGAAGGCTCACATTCTGCTGACGGCTGAATCATCTT	473
Db	663	CTGTGGGGGAGACGACACAAGTGAAGAAGCCCGTATAGCCATGGGTGAGGCAATATTT	722
QY	474	CATCATCACGGGCATGTGTGCTCATCCCTGTGAGCTGGGTTGCCAATGCCATCATCAG	533
Db	723	CATCGTGGCAGGCTTTGCCGCCCTTGTAAGCTTGTCTCTGTATGGCCATCAGATGTGCAC	782
QY	534	AGATTTCTATACTCAATAGTGAATGTGCCCAAAAACGTGAGCTTGAGAAGCTCTCTA	593
Db	783	AGACTTTTATAACCTTTGATCCCTACCAACATTAAGTATGAGTTTGGCCCTGCCACTT	842
QY	594	CTTAGGATGACCAACGGCACTGGTGCTGATTGTTGGAGGAGCTCTGTTCTGCTGCCTTT	653
Db	843	TATTGGCTGGGCAAGGCTCTGCCCTAGTCATCTCGGAGGTGCACCTGCTCTCCTGTTCTG	902
QY	654	TTGTTGCAACGAAAAGAGCAGTAGCTACAG	683
Db	903	TCCTGGGAATGAGAGCAAGGCTGGGTACCG	932

```

RESULT 6
US-09-489-847-93
; Sequence 93, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: PZ031P1
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/095,486
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/096,319
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 376
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 93
; LENGTH: 1722
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-489-847-93

```

Query Match	7.4%;	Score 149;	DB 4;	Length 1722;
Best-Local Similarity	55.3%;	Pred. No. 5,8e-29;		
Matches	306;	Conservative	2;	Mismatches 242; Indels 3; Gaps 1;
QY	97	ATGCCAACCCATGCGCTTAGAAATCGCTGGCTGTCTTCTGTGTGTGATGATGTGGGC	156	
Db	461	ATGGGGTCCGACGCGTTGGAGATCCTGGCCTGTGTCTGTGCTGTGGGCTGGGGGGGT	520	
QY	157	ACAGTGGCTGTCACTGTCAATGCTCCTCAGTGGAGAGCTGTGCGCCTTCAATTGAAAAACAACATC	216	
Db	521	CTGATCCTGGCGTGGCGGCTGCCCATGTGTGGCAAGTGACCGCCTTCTCGAACCAACAACATC	580	
QY	217	GTGCTTTTGAACAATTCTGGGAAGACTGTGATGAATTGCGTGAAGCAGGCTAACATC	276	
Db	581	GTGACGGCGCAACCAACCTGTGAAGGGGCTGTGTGATGTCTGCGCTGTGTCCAGACACSGGG	640	

QY 277 AGGATGACGTGCAAAATCTATGATTCCTGCTGCTCTTCTCCGAACTACAGGACCC 336
DB 641 CACATGCAGTGCAAAAGTGAACACTCGGTGCTGGCTCTGAGCACCGAGGTGACGGCGG 700
QY 337 AGAGACTGATGTGTGCTGCTCCGTGATGTCTTCTTGCTTTGATGATGGCCATCCTT 396
DB 701 CGGGCCTCACCCTGAGCGCGGTGCTGCTGGCGTTGTTGCGCTCTTCTGTAACCTGGCG 760
QY 397 GGCATGAATGCAACGAGTGCACGGGGACAAAGAGTGAAGGCTCACATTCTGCTG 456
DB 761 GGGCGCAGTGCACCACTGCG--GTGGCCCCGGGGCCCCGCAAGCGCGGTGGCGCTC 817
QY 457 ACGGCTGAATCATCTTCATCATCACGGGCATGTGTGCTCATCCCTGTGAGCTGGGTT 516
DB 818 ACGGAGGCGGTGCTTACTGTGTTTGGGGGCTGTGGCGCTCGTGCACCTGTGCTGCTTC 877
QY 517 GCCAATGCCATCATCAGAGATTCTATACTCATATAGTGAATGTGCCCAAAACGTGAG 576
DB 878 GCCAATGTGTGTCGCGGAGTTTACGACCCGCTGTGCCCCGTGTGCGAGAAATGAGAG 937
QY 577 CTTGAGAACTCTCTACTTAGATGAGCAACGCACTGTGTGATGTTGTTGAGAGCT 636
DB 938 CTGGGCGCAGCGCTGTACATCGGCTGGCGGCGCACCGCTGTCTCATGTAGCGCGCTGC 997
QY 637 CTGTTCTGCTGCG 649
DB 998 CTCTGTGCTGCG 1010

RESULT 7
US-09-886-683A-3
; Sequence 3, Application US/09886683A
; Patent No. 6627439
; GENERAL INFORMATION:
; APPLICANT: Hoevel, Thorsten
; APPLICANT: Koch, Stefan
; APPLICANT: Kubbies, Manfred
; APPLICANT: Mundigl, Olaf
; APPLICANT: Rueger, Petra
; TITLE OF INVENTION: Antibodies against SEM1 (p23)
; FILE REFERENCE: Case 20692
; CURRENT APPLICATION NUMBER: US/09/886,683A
; CURRENT FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: EP00113344.6
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: EP01107799.7
; PRIOR FILING DATE: 2001-04-05
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3443
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (221)..(853)
US-09-886-683A-3

Query Match 6.7%; Score 134.8; DB 4; Length 3443;
Best Local Similarity 52.5%; Pred. No. 4.2e-25;
Matches 295; Conservative 0; Mismatches 267; Indels 0; Gaps 0;
QY 86 CCCAGAGATATGCAACCATGCTTAGAAATGCTGGCTGTTTCTTGATGATG 145
DB 210 CCGAGCGATCATGCGCAACGGGGGCTGCAGCTGTGGCTTCACTTCGCTTCTGG 269
QY 146 GAATGTGGGCACAGTGTCTCACTGTCTATGCTCAGTGAGAGTGTGGCTTCAATG 205
DB 270 GATGATGGCGGCATCGTCAAGCACTGCCCTGCCCAAGTGAAGATTACTCTATGCCG 329
QY 206 AAACAACATCGTGTGTTTGAATACTCTGGAGAAGACTGTGATGAATTGCGTGAAGC 265

DB 330 GCGACAACATCGTAGCCGCCAGGCCATGTACGAGGGGCTGTGATGTCCTGCGTGC 389
QY 266 AGGTAACATCAGATGACGTGCAAAATCTATGATTCCTGCTGCTCTTCTCCGACC 325
DB 390 AGAGACCGGGAGATCCAGTGCAGAAAGTCTTGAATCTTGCTGTAATCTGACACACAT 449
QY 326 TACAGCAGCCAGAGACTGATGTGTGCTGCTTCCGTGATGTCCTTCTTGCTTTCATGA 385
DB 450 TGCAAGCAACCGGTGCTTGTATGTTGGTGGCATCTCTGGAGTGAATGCAATCTTGG 509
QY 386 TGGCCATCTTGGCATGAATGCAACAGGTGACAGGGGGAACAATGAGAAGTGAAGCTC 445
DB 510 TGGCCACCGTGGCATGAAGTGTATGAAGTCTTGGAAAGCATGAGGTGCAAGATGA 569
QY 446 ACATTCTGCTGACGGCTGGAATCATCTTCATCATCAAGGCATGTGTGCTCATCCCTG 505
DB 570 GGATGCTGTCAATGGGGGTGCGATATTTCTTTCAGGTCCTGCTATTTAGTTGCCA 629
QY 506 TGAGCTGGGTGCCAATGCCATCATCAGAGATTTCTATACTCAATAGTGAATGTGCC 565
DB 630 CAGCATGTAATGCAATAGATCGTTCAAGAATTCTATGACCTATGACCCAGTCAATG 689
QY 566 AAAACGTGACTTGGAGAAGCTCTTACTTAGATGGAACACGCACTGTGCTGATTTG 625
DB 690 CCAAGTAGAATTTGTCAGGCTCTTCTTCACTGGCTGGCTGCTGCTTCTCTGCTTC 749
QY 626 TTGAGGAGCTCTGTTGCTG 647
DB 750 TGGAGGTGCTTACTTGTCTG 771

RESULT 8
US-09-130-491-3
; Sequence 3, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for windows Version 3.0
; SEQ ID NO 3
; LENGTH: 3483
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (240)...(872)
US-09-130-491-3

Query Match 6.7%; Score 134.8; DB 4; Length 3483;
Best Local Similarity 52.5%; Pred. No. 4.2e-25;
Matches 295; Conservative 0; Mismatches 267; Indels 0; Gaps 0;
QY 86 CCCAGAGATATGCAACCATGCTTAGAAATGCTGGCTGTTTCTTGATGATG 145
DB 229 CCGAGCGATCATGCGCAACGGGGGCTGCAGCTGTGGCTTCACTTCGCTTCTGG 288
QY 146 GAATGTGGGCACAGTGTCTCACTGTCTATGCTCAGTGAGAGTGTGGCTTCAATG 205
DB 289 GATGATGGCGGCATCGTCAAGCACTGCCCTGCCCAAGTGAAGATTACTCTATGCCG 348
QY 206 AAACAACATCGTGTGTTTGAATACTCTGGAGAAGACTGTGATGAATTGCGTGAAGC 265
DB 349 GCGACAACATCGTAGCCGCCAGGCCATGTACGAGGGGCTGTGATGTCTGCTGCTGC 408

```
QY      266 AGGCTAACATCAGATGAGTGCAAAATCTATGATTCCTGCTGCTCTTTCTCCGAC 325
      |||
Db      409 AGAGACCGGGCAGATCCAGTCAAAAGTCTTTGACTCCTTGTAATCTGAGCAGACAT 468
      |||
QY      326 TACAGGACCCAGAGACTGATGTGTGCTGCTTCGGTATGCTCTTCTTGCTTCATGA 385
      |||
Db      469 TGCAGGACCCCGTGCCTTATGTGTGTGGCATCTCTGGAGTATAGCAATCTTTG 528
      |||
QY      386 TGGCCATCCTTGGCATGAATGACCAAGTGCACGGGGACAAATGAGAAGTGAGGCTC 445
      |||
Db      529 TGGCCACCGTTGGCATGAGTATGAGTGCTTGAAGACGATGAGGTGACAGAGATGA 588
      |||
QY      446 ACATTCTGCTGACGCGCTGGAATCATCTTCATCATCAAGGCATGGTGTGCTCATCCCTG 505
      |||
Db      589 GGATGCTGTTCATTTGGGGTGCAGATATTTCTTGTGACAGTCTGGCTATTTAGTTGCCA 648
      |||
QY      506 TGAGCTGGCTGCCAATGCCATCATCAGAGATTTCTATACTCAATAGTGAATGTGCC 565
      |||
Db      649 CAGCATGGTATGCAATAGATCGTTCAAGATTTCTATGACCCCTATGACCCAGTCAATG 708
      |||
QY      566 AAAACGTGAGCTTGAGAGAGCTCTCTACTTAGATGAGCAGCAGCACTGTGCTGATG 625
      |||
Db      709 CCAGGTACGAATTTGGTCAAGCTCTCTTCACTGAGCTGGGCTGCTCTTCTCTGCTTC 768
      |||
QY      626 TTGAGAGAGCTCTGTTCTGCTG 647
      |||
Db      769 TGGAGGTGCCCTACTTGTCTG 790
      |||
```

RESULT 9

```
US-09-489-847-125
; Sequence 125, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: P2031P1
; CURRENT APPLICATION NUMBER: US/09/489,847
; EARLIER FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/095,486
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/096,319
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 376
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 125
; LENGTH: 1380
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-489-847-125
```

```
Query Match      6.4%; Score 129; DB 4; Length 1380;
Best Local Similarity 55.7%; Pred. No. 8.3e-24;
Matches 308; Conservative 0; Mismatches 240; Indels 5; Gaps 3;
```

```
QY      97 ATGCAACCATGCTTAGAAATCGCTGGCTGTTTCTTGTTGTTGGAATGTGGGC 156
      |||
Db      135 ATGGGTCGCGACGCTTGAGATCTGGCTGTGTGCTGTGGCTGGGCGGT 194
      |||
QY      157 ACAGTGGCTGCACTGTCAAGCTCAGTGAGAGTGTGGCTTCAATTGAAAAACAATC 216
      |||
Db      195 CTGATCCTGGCGTGGCGCTGCCATGTGACAGGTGACCGCTTCTTGACCAACAATC 254
      |||
QY      217 GTGTTTGAATACTTCTGGGAAGACTGTGATGATTGCGTGAAGCAGGCTAATC 276
      |||
```

```
Db      255 GTGACGGCGCAGACCACCTGGAAAGGGCTGTGATGTCGCTGCTGTGACAGACCGGC 314
      |||
QY      277 AGATGACGTGCAAAATCTATGATTCCTGCTGCTCTTTCTCCGACCTACAGGACGC 336
      |||
Db      315 A-CATGACGTCAAAAGTGTACGACTCGGTGTGCTCTGTGACACCGAGGTGACAGCGCG 373
      |||
QY      337 AGAGACTGATGTGTGCTGCTTCCGTATGATGTCCTTCTTGCTTCATGATGCCATCTT 396
      |||
Db      374 CGGGCGCTCACCGTGAGCGCCGTGCTGCTGCGCTTGTGCGCTTCTTGACCCCTGGCG 433
      |||
QY      397 GGCATGAATGACACAGGTGACCGGGGACAAATGAGAAGTGAGGCTCACATTTGCTG 456
      |||
Db      434 GCGCGCAGTGCACACCTGC--GTGGCCCGGGGCCGCAAGGCGGTGTGGCCCTC 490
      |||
QY      457 ACGCTGAATCATCTTTCATCATCAAGGCATGGTGTGCTCATCCCTGTGAGCTGGCTT 516
      |||
Db      491 ACGGAGCGGTCTTACCTGTTTTCGGGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCT 550
      |||
QY      517 GCCAATGCATCATCAGAGATTCTATACTCAATAGTGAATGTGCCCCAAAAACGTGAG 576
      |||
Db      551 GCCAATGTTGCTCCGCGAGTTTACGACCCGCTGTGCCCCGTGCGAGAGTACGAG 610
      |||
QY      577 CTTGAGAGCTCTTACTTAGATGAGCAGCAGCACTGTGCTGATTGTTGAGAGACT 636
      |||
Db      611 CTGGCGCA-CGCTGTACATCGGCTGGCGGCGCACCGCGCTGTGATGAGCGGCTGC 669
      |||
QY      637 CTGTTCTGCTGCG 649
      |||
Db      670 CTCTTGTGCTGCG 682
      |||
```

RESULT 10

```
US-09-663-600A-45
; Sequence 45, Application US/09663600A
; Patent No. 6573068
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclerc, Aymeric
; APPLICANT: Bouquelere, Lydie
; TITLE OF INVENTION: EXTENDED CDNAS FOR SECRETED PROTEINS
; FILE REFERENCE: 31.US3.CIP
; CURRENT APPLICATION NUMBER: US/09/663,600A
; CURRENT FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/191,997
; PRIOR FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: 60/066,677
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/069,957
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/074,121
; PRIOR FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: 60/081,563
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/096,116
; PRIOR FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: 60/099,273
; PRIOR FILING DATE: 1998-09-04
; NUMBER OF SEQ ID NOS: 229
; SOFTWARE: Patent.pm
; SEQ ID NO 45
; LENGTH: 1524
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 160..231
OTHER INFORMATION: Von Heijne matrix
OTHER INFORMATION: score 5.6999980926514
NAME/KEY: polyA_signal
LOCATION: 1510..1515
NAME/KEY: polyA_site
LOCATION: 1506..1519
```


QY	277	AGGATGCA	GTGCAAAATCTATGATTCCTGCTGCTCTTTCTCCGCACTACAGGACGCC	336
Db	216	ATCACCCA	GTGTGACATCTATAGACCCCTTCTGGGCTGCCCGCTGACATCCAGGCTGCC	275
QY	337	AGAGACTG	ATGTGTGCTGCTTCCGTGATGTCTTCTTGGCTTTCATGATGCCATCTT	396
Db	276	CAGGCATG	ATGGTGCATCCAGTGCATCTCTCCCTGGCTGCATTATCTCTGTGTG	335
QY	397	GGCATGAA	TGCCACAGGTGCACGGGGACAATGAGAAAGTGAAAGGCTCACATTTCTG	456
Db	336	GGCATGAG	ATGCACAGTCTTCTGCCAGAAATCC--CGAGCCAAGAAGAGAGTGGCGTA	392
QY	457	ACGCTGGA	TCACTTTCATCATCAGGGCATGTGTGTCTCATCCCTGTGAGCTGGTT	516
Db	393	GCAGGTGA	GTCTTTTCATCTTGAGGCTCTCGGATTCATTCCTGTGCTGGAAT	452
QY	517	GCCAATGC	ATCATCAGAGATTTCTAATCAATAGTAGTAATGTGCCAAAAACGTGAG	576
Db	453	CTTCATGG	ATCCTACGGGACTTCTACTCACCACTGCTGCCAGACAGCATGAATTTGAG	512
QY	577	CTTGAGAG	CTCTCTACTTAGATGACCAACGGCACTGGTCTGATTGTTGAGAGACT	636
Db	513	ATTGAGAG	AGGCTCTTACTTTGGGCATTATTCTTCCCTGTTCTCCCTGATAGCTGAAATC	572
QY	637	CTGTTCGT	GCGTTTTTGTGTGCACGAAAGAGCAGTAGCTACAGATTACTCGATACT	696
Db	573	ATCCTCTG	CTTTTCTCTGCTCATCCAGAGAAATCGCTCCAACTACTACGATGCCCTACCA	632
QY	697	TCCCATGC	ACAACCCAAAAAGTTATC 724	
Db	633	GCCCAAC	CTTTGGCCACAAGAGCTCTC 660	

```

RESULT 12
US-09-188-930-55
; Sequence 55, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 55
; . LENGTH: 413
; . TYPE: DNA
; . ORGANISM: Human
US-09-188-930-55

```

Query Match	4.7%;	Score 94.8;	DB 3;	Length 413;
Best Local Similarity	54.6%;	Pred. No. 3.6e-15;		
Matches 189; Conservative	0;	Mismatches 157;	Indels 0;	Gaps 0;

[illegible]

Db 201 ATCGTGGCAGGTCCTTGCCGCGCTTGCTAGCTTGCTCCTGGTAAGCCATCAGATTGTACCA 260

QY 535 GATTTCATACTCATAGTGAATGTTGCCAAAAACGTGAGCTTGAGAACTCTCTAC 594

Db 261 GACTTTTATAACCCCTTTGATCCCTAACCAACATTAAAGTATGAGTTTGCCCGCTGCCATCTTT 320

QY 595 TTAGATGACCAACGCGCACTGCTGATTTGTTGAGAGACTCTGT 640

Db 321 ATTGGCTGGGCAAGGTCGTGCCCTAGTCATCCTGGAGGTGCACCTGT 366

```

RESULT 13
US-09-312-283C-55
; Sequence 55, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Orust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated
; FILE REFERENCE: 11000.1011c2
; CURRENT APPLICATION NUMBER: US/09/312,283C
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 425
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 55
; LENGTH: 413
; TYPE: DNA
; ORGANISM: Mouse
US-09-312-283C-55

```

Query Match	4.7%;	Score 94.8;	DB 4;	Length 413;
Best Local Similarity	54.6%;	Pred. No. 3.6e-15;		
Matches 189;	Conservative	0;	Mismatches 157;	Indels 0;
				Gaps 0;

QY	295	TATGATTCCCTGCTGGCTCTTTTCTCCCGAACCCTACAGGCAGGCCAGAGACTGATGTGCT	354
Db	21	TACGACTCGGTGCTCGCCCTGTCCGGGCTTGCAAGGCCACTTCGAGCCCTAATGGTGGTC	80
QY	355	GCTTCCGATGATGCCCTTCTTGGCTTTCATGATGGCCATCCTTGGCATGAATGCACCAGG	414
Db	81	TCCCTGGTGTGGGCTTCCCTGGCCATGTTTGTGGCCACGATGGGCATGAAGTGCACGGCC	140
QY	415	TGCACGGGGACAATGAGAAGGTGAAGGCTCACATTCTGTGACGGCTGGAATCATCTTC	474
Db	141	TGTGGGGAGACGACAAAGTGAAGAGCCCCGATATAGCCATGGGTGAGGCATAATTTTC	200
QY	475	ATCATCACGGGCATGTGTGCTCATCCCTGTGAGCTGGGTGCCAATGCCATCATCAAGA	534
Db	201	ATCGTGGCAGGTCTTGCCGCTTGTGTAGCTTGCTCCTGGTATGGCCATCAGATTGTCA	260
QY	535	GATTTCATTACTCAATAGTAATGTTGCCCAAAAACGTGAAGCTTGAGAAGCTCTAC	594
Db	261	GACTTTATAACCCCTTTGATCCCTACCAACATTAAGTATGAGTTTGGCCCCTGCCATCTT	320
QY	595	TTAGATGACCAACGGCAGCTGTGCTGATTTGTTGAGGAGCTCTGT	640
Db	321	ATTGGCTGGGCAAGGCTTGCCCTAGTCATCCTGGGAGGTGCACTGT	366

RESULT 14
US-09-404-879A-104
? Sequence 104, Application US/09404879A
? Patent No. 6468546
? GENERAL INFORMATION:
? APPLICANT: Mitcham, Jennifer L.
? APPLICANT: King, Gordon E.
? APPLICANT: Algate, Paul A.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 104
; LENGTH: 441
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-404-879A-104

Query Match 4.5%; Score 89.8; DB 4; Length 441;
Best Local Similarity 59.7%; Pred. No. 7.4e-14;
Matches 151; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 92 GGATAATGGCAACCCATGCTTAGAAATCGCTGGCTGTTCTTGTTGGTGAATGG 151
Db 178 GAACAATGGCTCCATAGGGCTACAGTAAATGGGCATCGCGCTGGCCGCTCGGCTGGC 237
QY 152 TGGGCACAGTGGCTGTCACTGTCACTGACCTCAGTGGAGAGTGTGGCCTTCATTGAAACA 211
Db 238 TGGCCGTATGCTGTGTGCTGCGCGCTGCCCATGTGGCGCGGTGACGGCCTTCATCGGCAGCA 297
QY 212 ACATCGTGTGTTTGAATACTTCTGGGAAGACTGTGATGAATTGCGTGAGGCAGGCTA 271
Db 298 ACATGTGACCTCGCAGACCATCTGGAGGGCCTATGATGAATCGTGTGAGAGCA 357
QY 272 ACATCAGATGCAAGTGAATACTATGATTCCTGCTGCTCTTTCTCCGACCTACAGG 331
Db 358 CCGGCCAGATGCAAGTGAAGTGTACGACTCGCTGCGACTGCGCAGAGCCTGAGG 417
QY 332 CAGCCAGAGGACT 344
Db 418 CGGCCCGGCCCT 430

RESULT 15
US-09-938-933-104
; Sequence 104, Application US/09338933
; Patent No. 6488931
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
; TITLE OF INVENTION: OVARIAN CANCER
; FILE REFERENCE: 210121.462C1
; CURRENT APPLICATION NUMBER: US/09/338,933
; CURRENT FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 104
; LENGTH: 441
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-938-933-104

Query Match 4.5%; Score 89.8; DB 4; Length 441;
Best Local Similarity 59.7%; Pred. No. 7.4e-14;
Matches 151; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 92 GGATAATGGCAACCCATGCTTAGAAATCGCTGGCTGTTCTTGTTGGTGAATGG 151
Db 178 GAACAATGGCTCCATAGGGCTACAGTAAATGGGCATCGCGCTGGCCGCTCGGCTGGC 237
QY 152 TGGGCACAGTGGCTGTCACTGTCACTGACCTCAGTGGAGAGTGTGGCCTTCATTGAAACA 211
Db 238 TGGCCGTATGCTGTGTGCTGCGCGCTGCCCATGTGGCGCGGTGACGGCCTTCATCGGCAGCA 297
QY 212 ACATCGTGTGTTTGAATACTTCTGGGAAGACTGTGATGAATTGCGTGAGGCAGGCTA 271
Db 298 ACATGTGACCTCGCAGACCATCTGGAGGGCCTATGATGAATCGTGTGAGAGCA 357

QY 272 ACATCAGATGCAAGTGAATACTATGATTCCTGCTGCTGCTCTTTCTCCGACCTACAGG 331
Db 358 CCGGCCAGATGCAAGTGAAGTGTACGACTCGCTGCTGCGACTGCGCAGAGCCTGAGG 417
QY 332 CAGCCAGAGGACT 344
Db 418 CGGCCCGGCCCT 430

Search completed: September 2, 2004, 23:34:11
Job time : 114 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2004, 21:14:05 / Search time 5283 Seconds
(without alignments)
16490.535 Million cell updates/sec

Title: US-10-063-732-119
Perfect score: 2010
Sequence: 1 ggaactgtctctctctctg.....aattgtacatttttctaatt 2010

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: gb_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2010	100.0	2010	6	AX092388 Sequence
2	2010	100.0	2010	6	AX376290 Sequence
3	2010	100.0	2010	6	AX697259 Sequence
4	2010	100.0	2010	9	AY358707 Homo sapi
5	2008.4	99.9	137955	9	AP000884 Homo sapi
6	2008.4	99.9	340000	9	AP001707 Homo sapi
7	1944	96.7	267172	9	BS000177 Pan trogl
8	1935	96.3	176580	9	BS000178 Pan trogl
9	1929.4	96.0	1931	9	HS4250711 Homo sapi
10	1802	89.7	1890	6	AX879004 Sequence
11	1802	89.7	1890	6	BD157589 Primer fo
12	1802	89.7	1890	9	AX022269 Homo sapi
13	1675.8	83.4	1700	9	BC058004 Homo sapi
14	1550.4	77.1	1835	9	BC020866 Homo sapi
15	1352.8	67.3	191923	2	AP001846 Homo sapi
16	1054.8	52.5	1090	9	HSMB00324
17	950	47.3	1002	6	AX497202 Sequence
18	902.2	44.9	81583	2	AC131927 Homo sapi
19	845.6	42.1	240479	2	AC096377 Rattus no
20	789.2	39.3	2325	10	BC003868 Mus muscu
21	784.6	39.0	21278	2	AC113180 Mus muscu
22	784.6	39.0	256720	10	AC110241 Mus muscu
23	761	37.9	837	6	AX866683 Sequence
24	761	37.9	837	6	BD146745 Primer fo
25	535.8	26.7	577	6	AX872185 Sequence
26	535.8	26.7	577	6	BD152247 Primer fo
27	494	24.6	678	6	E31609
28	494	24.6	678	10	AF087826 Mus muscu
29	493.6	24.6	191923	2	AP001846 Homo sapi
30	493.2	24.5	615	6	AR414198 Sequence
31	493.2	24.5	615	6	BD109751 EST and e
32	365.4	18.2	429	6	AX203264 Sequence
33	308.2	15.3	368	6	BD204000 5'EST and
34	277.6	13.8	675	9	HS4250712 Homo sapi
35	277.2	13.8	833	6	AX497204 Sequence
36	264.6	13.2	1160	6	AX697263 Sequence
37	264.6	13.2	1160	9	AY358094 Homo sapi
38	210.4	10.5	1449	10	BC058186 Mus muscu
39	210.4	10.5	142656	2	AC133582 Mus muscu
40	210.4	10.5	220013	10	AC122821 Mus muscu
41	209.4	10.4	1167	9	HS4130941 Homo sapi
42	209.4	10.4	1231	6	AX058215 Sequence
43	209.4	10.4	2050	9	AK091002 Homo sapi
44	209.4	10.4	43411	9	AC004643 Homo sapi
45	209.4	10.4	180354	9	AC108134 Homo sapi

ALIGNMENTS

RESULT 1
AX092388
LOCUS AX092388 2010 bp DNA linear PAT 21-MAR-2001
DEFINITION Sequence 119 from Patent WO0116318.
ACCESSION AX092388
VERSION AX092388.1 GI:13444508
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Eaton,D.L., Filvaroff,E., Gerritsen,M.E., Goddard,A.,
Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and
Wood,W.I.

TITLE										Secreted and transmembranepolypeptides and nucleic acids encoding the same									
JOURNAL										Patent: WO 0116318-A 119 08-MAR-2001; Genentech, Inc. (US)									
FEATURES										Location/Qualifiers									
source										1..2010 /organism="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606"									
ORIGIN																			
Query Match										100.0%; Score 2010; DB 6; Length 2010;									
Best Local Similarity										100.0%; Pred. No. 0;									
Matches 2010; Conservative										0; Mismatches 0; Indels 0; Gaps 0;									
OY	1	GGAAAACTGTTCTCTTGTGGCACA	GAGAACCCCTGCTTCAAGCAGAA	GTAGAGTTC	60														
Db	1	GGAAAACTGTTCTCTTGTGGCACA	GAGAACCCCTGCTTCAAGCAGAA	GTAGAGTTC	60														
OY	61	CGAGTCCAGCTGGCTAAACTCAT	CCAGAGATATGGCAACCCATGCTTAGAAATC	120															
Db	61	CGAGTCCAGCTGGCTAAACTCAT	CCAGAGATATGGCAACCCATGCTTAGAAATC	120															
OY	121	GCTGGGCTGTTCTTGGTGTGTAAT	GTGGGCACAGTGGCTGTCACTGTGACCT	180															
Db	121	GCTGGGCTGTTCTTGGTGTGTAAT	GTGGGCACAGTGGCTGTCACTGTGACCT	180															
OY	181	CAGTGGAGAGTGTGGGCTTCATT	GAAAAACAACATCGTGTGTTTGAACCTTCGGAA	240															
Db	181	CAGTGGAGAGTGTGGGCTTCATT	GAAAAACAACATCGTGTGTTTGAACCTTCGGAA	240															
OY	241	GGACTGTGATGAATTGCGTGAAG	CGCTAACATCAGATGCAAGTGCAAAAATCTATGAT	300															
Db	241	GGACTGTGATGAATTGCGTGAAG	CGCTAACATCAGATGCAAGTGCAAAAATCTATGAT	300															
OY	301	TCCCTGTGGCTCTTCTCCGGA	ACCTACAGGACCCAGAGACTGATGTGCTGCTCC	360															
Db	301	TCCCTGTGGCTCTTCTCCGGA	ACCTACAGGACCCAGAGACTGATGTGCTGCTCC	360															
OY	361	GTGATGTCTTCTTGGCTTTCAT	GATGGCCATCTTGGCATGAATGCACCAAGTGCACG	420															
Db	361	GTGATGTCTTCTTGGCTTTCAT	GATGGCCATCTTGGCATGAATGCACCAAGTGCACG	420															
OY	421	GGGACAATGAGAAGGTGAAGG	CTCACATTGTCTGACGCGCTGGAATCATCTTCATC	480															
Db	421	GGGACAATGAGAAGGTGAAGG	CTCACATTGTCTGACGCGCTGGAATCATCTTCATC	480															
OY	481	ACGGGCATGTGTGCTCATCC	CTGTGAGCTGGGTTGCCAATGCCATCATCAGAGATTTC	540															
Db	481	ACGGGCATGTGTGCTCATCC	CTGTGAGCTGGGTTGCCAATGCCATCATCAGAGATTTC	540															
OY	541	TATACTCAATAGTAATGTG	CCCCAAAAACGTGAGCTTGAGAGAGCTCTCTACTTAAGA	600															
Db	541	TATACTCAATAGTAATGTG	CCCCAAAAACGTGAGCTTGAGAGAGCTCTCTACTTAAGA	600															
OY	601	TGACCAACGGCACTGTGTG	ATGTTGTGAGAGAGCTGTGTCGCTGCTTTTGTGTC	660															
Db	601	TGACCAACGGCACTGTGTG	ATGTTGTGAGAGAGCTGTGTCGCTGCTTTTGTGTC	660															
OY	661	AACGAAAAGACAGTAGCTAC	AGATACTCGATACCTTCCCATCGCACAACCCAAAAAGT	720															
Db	661	AACGAAAAGACAGTAGCTAC	AGATACTCGATACCTTCCCATCGCACAACCCAAAAAGT	720															
OY	721	TATCACACCGGAAAAGAAG	TCAACCGAGCGTCTACTCCAGAAGTCAGTATGTAGTGTGT	780															
Db	721	TATCACACCGGAAAAGAAG	TCAACCGAGCGTCTACTCCAGAAGTCAGTATGTAGTGTGT	780															
OY	781	ATGTTTTTTTAACCTTAATA	AGCCATGCAAAATGACMAAAATCTATTTACTTTCTCA	840															
Db	781	ATGTTTTTTTAACCTTAATA	AGCCATGCAAAATGACMAAAATCTATTTACTTTCTCA	840															
OY	841	AAATGACCCCAAAAGAACT	TGATTACTGTCTTAAGCTTAATCTTAATTACAGGA	900															
Db	841	AAATGACCCCAAAAGAACT	TGATTACTGTCTTAAGCTTAATCTTAATTACAGGA	900															

Db	841	AAATGACCCCAAGAAACTTGGATTACTGTTCCTTAACCTGCTTAATCTTAATTAACAGGA	900
OY	901	ACTGTGATCAGCTATTTATGATTTCTATTAAGCTATTTCAGCAGAAATGAGATATTTAAACC	960
Db	901	ACTGTGATCAGCTATTTATGATTTCTATTAAGCTATTTCAGCAGAAATGAGATATTTAAACC	960
OY	961	AATGCTTGATTGTTCTAGAAAAGTATAGTAATTTGTTTTCTAAGGTGTTCAAGCATCTA	1020
Db	961	AATGCTTGATTGTTCTAGAAAAGTATAGTAATTTGTTTTCTAAGGTGTTCAAGCATCTA	1020
OY	1021	CTCTTTTATCATTTACTTCAAAATGACATTGCTAAAGACTGCATTTATTTACTACTGTA	1080
Db	1021	CTCTTTTATCATTTACTTCAAAATGACATTGCTAAAGACTGCATTTATTTACTACTGTA	1080
OY	1081	ATTTCTCCACGACATAGCATTTATGTACATAGATGAGTGTAAACATTTATCTCACATAGA	1140
Db	1081	ATTTCTCCACGACATAGCATTTATGTACATAGATGAGTGTAAACATTTATCTCACATAGA	1140
OY	1141	GACATGCTTATATGTTTATTTAAATGAATGCCAGTCCATTACACTGAATAAATAGA	1200
Db	1141	GACATGCTTATATGTTTATTTAAATGAATGCCAGTCCATTACACTGAATAAATAGA	1200
OY	1201	ACTCAACTATGCTTTTCAGGGAATCATGATAGGTTGAAGAAGTTACTATTAATTG	1260
Db	1201	ACTCAACTATGCTTTTCAGGGAATCATGATAGGTTGAAGAAGTTACTATTAATTG	1260
OY	1261	TTTAAACAGCTTAGGGATTAATGCTCCTCCATTTATGAAGATTAAATGAAGCTT	1320
Db	1261	TTTAAACAGCTTAGGGATTAATGCTCCTCCATTTATGAAGATTAAATGAAGCTT	1320
OY	1321	TAATCAGACTTGTAAAGGAATGAATGGCTTCTGATATGCTGTTTTTAGCCTAGGAG	1380
Db	1321	TAATCAGACTTGTAAAGGAATGAATGGCTTCTGATATGCTGTTTTTAGCCTAGGAG	1380
OY	1381	TTAGAAATCCTAATCTTCTTATCCTCTCTCCAGAGGCTTTTTTCTGTGTATTAA	1440
Db	1381	TTAGAAATCCTAATCTTCTTATCCTCTCTCCAGAGGCTTTTTTCTGTGTATTAA	1440
OY	1441	ATTAACATTTTAAAAACGACATATTTGTCAAGGGCTTTGCATTCAAACTGCTTTCC	1500
Db	1441	ATTAACATTTTAAAAACGACATATTTGTCAAGGGCTTTGCATTCAAACTGCTTTCC	1500
OY	1501	AGGCTATATCTCAGAAAGAAAGATAAAGTGTGATCTAAGAAAAAGTATGTTTAGGAA	1560
Db	1501	AGGCTATATCTCAGAAAGAAAGATAAAGTGTGATCTAAGAAAAAGTATGTTTAGGAA	1560
OY	1561	AGTGAATAATTTTTGTTTTGTATTTGAAGAAGATGATGATTTTGACAAAGAAATCAT	1620
Db	1561	AGTGAATAATTTTTGTTTTGTATTTGAAGAAGATGATGATTTTGACAAAGAAATCAT	1620
OY	1621	ATATGTATGATATATTTTAAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATATA	1680
Db	1621	ATATGTATGATATATTTTAAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATATA	1680
OY	1681	AATAAAGACAGAAAAATATGTCTTGTTTTTCATTGCTTACCAAAAAACAACAAGAA	1740
Db	1681	AATAAAGACAGAAAAATATGTCTTGTTTTTCATTGCTTACCAAAAAACAACAAGAA	1740
OY	1741	AAAAAGTTGCTTTGAGAACTTCACTGCTCTATGTGGGTACCTGAGTCAAAATTGTC	1800
Db	1741	AAAAAGTTGCTTTGAGAACTTCACTGCTCTATGTGGGTACCTGAGTCAAAATTGTC	1800
OY	1801	ATTTTGTCTGTGAAAAATAATTCTCTTCTGTACCATTTCTGTTTAGTTTACTTAA	1860
Db	1801	ATTTTGTCTGTGAAAAATAATTCTCTTCTGTACCATTTCTGTTTAGTTTACTTAA	1860
OY	1861	ATCTGTAAATACTGTATTTTCTGTTTATTCCAAATTTGATGAAACTGACAATCCAAATT	1920
Db	1861	ATCTGTAAATACTGTATTTTCTGTTTATTCCAAATTTGATGAAACTGACAATCCAAATT	1920
OY	1921	GAAAGTTTGTGTCGACGCTGTCTAGCTTAAATGAATGTGTTCTATTTGCTTTATACATT	1980
Db	1921	GAAAGTTTGTGTCGACGCTGTCTAGCTTAAATGAATGTGTTCTATTTGCTTTATACATT	1980

Qy 1981 TATATTAAATGTACATTTTCTAATT 2010
Db 1981 TATATTAAATGTACATTTTCTAATT 2010

RESULT 2
AX376290
LOCUS AX376290 2010 bp DNA linear PAT 01-MAR-2002
DEFINITION Sequence 357 from Patent WO0168848.
ACCESSION AX376290
VERSION AX376290.1 GI:19170536
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Baker,K.P., Chen,J., Desnoyers,L., Goddard,A., Godowski,P.J.,
Gurney,A.L., Pan,J., Smith,V., Watanabe,C.K., Wood,W.I. and
Zhang,Z.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0168848-A 357 20-SEP-2001;
Genentech, Inc. (US)
FEATURES
source 1..2010
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 2010; DB 6; Length 2010;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2010; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGAATAACTGTTCTTCTGTGGCAGAGAAACCCTGCTTCAAAGCAGAGTACAGTTC 60
Db 1 GGAATAACTGTTCTTCTGTGGCAGAGAAACCCTGCTTCAAAGCAGAGTACAGTTC 60
Qy 61 CGGAGTCCAGCTGGCTAAACTCATCCCGAGGATATGGCAACCCTTGAAGATC 120
Db 61 CGGAGTCCAGCTGGCTAAACTCATCCCGAGGATATGGCAACCCTTGAAGATC 120
Qy 121 GCTGGGCTGTTCTTGGTGGTGTGAATGGTGGGCAAGTGGCTGTCACTGTGCTT 180
Db 121 GCTGGGCTGTTCTTGGTGGTGTGAATGGTGGGCAAGTGGCTGTCACTGTGCTT 180
Qy 181 CAGTGGAGAGTGTGGGCTTCATTTGAAAACAACATCGTGTGTTTGAACCTTGGGA 240
Db 181 CAGTGGAGAGTGTGGGCTTCATTTGAAAACAACATCGTGTGTTTGAACCTTGGGA 240
Qy 241 GGAAGTGGATGAATGGTGGTGGGAGGCTAAACATCAGATGCAAGTGAATATGAT 300
Db 241 GGAAGTGGATGAATGGTGGTGGGAGGCTAAACATCAGATGCAAGTGAATATGAT 300
Qy 301 TCCCTGCTGCTCTTCTCCGGAACCTACAGGAGGAGGAGGAGTGTGTGCTTCC 360
Db 301 TCCCTGCTGCTCTTCTCCGGAACCTACAGGAGGAGGAGGAGTGTGTGCTTCC 360
Qy 361 GTGATGTCCTTCTTGGCTTTCATGATGGCCATCCTTGGCATGAATGCAACGATGCA 420
Db 361 GTGATGTCCTTCTTGGCTTTCATGATGGCCATCCTTGGCATGAATGCAACGATGCA 420
Qy 421 GGGGACAATGAGAGGAGGAGGCTCACATTTCTGTGACGGCTGGAATCATCTTCATC 480
Db 421 GGGGACAATGAGAGGAGGAGGCTCACATTTCTGTGACGGCTGGAATCATCTTCATC 480
Qy 481 ACGGCGATGTGTGCTCATCCCTGTGAGCTGGGTGCCAATGCCATCATCAGATTTTC 540
Db 481 ACGGCGATGTGTGCTCATCCCTGTGAGCTGGGTGCCAATGCCATCATCAGATTTTC 540
Qy 541 TATTAATCAATAGTGAATGTTGCCCAAAAAGTGAAGTGTGAGAGAGCTCTTACTTAGA 600

Db 541 TATTAATCAATAGTGAATGTTGCCCAAAAAGTGAAGTGTGAGAGAGCTCTTACTTAGA 600
Qy 601 TGGACCAAGGCACTGTGTGATGTTGGAGAGCTCTGTTCTGTGCTGTTTGTGTC 660
Db 601 TGGACCAAGGCACTGTGTGATGTTGGAGAGCTCTGTTCTGTGCTGTTTGTGTC 660
Qy 661 AACGAAAGAGCAGTAGCTACAGTACTCGATACCTTCCATCGCAGACCAAAAAAGT 720
Db 661 AACGAAAGAGCAGTAGCTACAGTACTCGATACCTTCCATCGCAGACCAAAAAAGT 720
Qy 721 TATCACACCGGAAAGAGTCAACCGAGCTCTACTCCAGAGTCAAGTATGTGTGTGT 780
Db 721 TATCACACCGGAAAGAGTCAACCGAGCTCTACTCCAGAGTCAAGTATGTGTGTGT 780
Qy 781 ATGTTTTTTAACTTTACTATTAAGCCATGCAAAATGACAAAATCTATATTACTTCTCA 840
Db 781 ATGTTTTTTAACTTTACTATTAAGCCATGCAAAATGACAAAATCTATATTACTTCTCA 840
Qy 841 AAATGAGCCCCAAGAAACTTTGATTTACTGTTCTTAACTGCTTAATCTTAATACAGA 900
Db 841 AAATGAGCCCCAAGAAACTTTGATTTACTGTTCTTAACTGCTTAATCTTAATACAGA 900
Qy 901 ACTGTGATCAGCTATTATATATTTCTATTAAGCTATTTCAGCAGATGAGATATTAACCC 960
Db 901 ACTGTGATCAGCTATTATATATTTCTATTAAGCTATTTCAGCAGATGAGATATTAACCC 960
Qy 961 AATGCTTTGATGTTCTTAGAAAGTATAGTAATTTGTTTCTAAGTGGTCAAGCATCTA 1020
Db 961 AATGCTTTGATGTTCTTAGAAAGTATAGTAATTTGTTTCTAAGTGGTCAAGCATCTA 1020
Qy 1021 CTCTTTTATCAATTTACTTCAAAATGACATTTGCTAAAGACTGCATATTTTTACTACTGTA 1080
Db 1021 CTCTTTTATCAATTTACTTCAAAATGACATTTGCTAAAGACTGCATATTTTTACTACTGTA 1080
Qy 1081 ATTTCGCAAGACATAGCATATGTAATAGTGAAGTGAACATTTATATCTCACATAGA 1140
Db 1081 ATTTCGCAAGACATAGCATATGTAATAGTGAAGTGAACATTTATATCTCACATAGA 1140
Qy 1141 GACATGCTATATGTTTAAATGAATGAATGCCAGTCCATTACCTGAATTAATAGA 1200
Db 1141 GACATGCTATATGTTTAAATGAATGAATGCCAGTCCATTACCTGAATTAATAGA 1200
Qy 1201 ACTCAATATGCTTTTCAAGGAATCATGATAGGCTTGAAGAGGTTACTATTAATG 1260
Db 1201 ACTCAATATGCTTTTCAAGGAATCATGATAGGCTTGAAGAGGTTACTATTAATG 1260
Qy 1261 TTTAAACAAGCTTAGGATTAATGCTCCTCCATTTATATGAAGATTAATAAGAGCTT 1320
Db 1261 TTTAAACAAGCTTAGGATTAATGCTCCTCCATTTATATGAAGATTAATAAGAGCTT 1320
Qy 1321 TAATCAGATTTGAAGAAATGAATGGCTTTCTGATATGCTGTTTTTAAAGCTAGAG 1380
Db 1321 TAATCAGATTTGAAGAAATGAATGGCTTTCTGATATGCTGTTTTTAAAGCTAGAG 1380
Qy 1381 TTGAATTCCTAATCTTATATCCTCTTCTCCAGAGGCTTTTCTTCTGTGTATTA 1440
Db 1381 TTGAATTCCTAATCTTATATCCTCTTCTCCAGAGGCTTTTCTTCTGTGTATTA 1440
Qy 1441 ATTAACATTTTAAACGAGATATTTGTCAAGGGCTTTGCAATCAAACTGCTTTCC 1500
Db 1441 ATTAACATTTTAAACGAGATATTTGTCAAGGGCTTTGCAATCAAACTGCTTTCC 1500
Qy 1501 AGGCTATATCTCAGAAAGAGATAAAGTGTGATCTAAGAAAAAGTGTGTTAGAA 1560
Db 1501 AGGCTATATCTCAGAAAGAGATAAAGTGTGATCTAAGAAAAAGTGTGTTAGAA 1560
Qy 1561 AGTGAATAATTTTGTGTTTGTATTTGAAGAAATGATGATTTGAACAAGAAATCAT 1620
Db 1561 AGTGAATAATTTTGTGTTTGTATTTGAAGAAATGATGATTTGAACAAGAAATCAT 1620
Qy 1621 ATATGTATGATATATTTTAATGAATGATTTGAGTACAGACTTTGAGGTTTCATCAATTA 1680

Db	1621	ATATGATGATATATATTTTAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATATA	1680
QY	1681	AATATAAGACAGAAAAATATGTCTTGGTTTTCATTTGTCTTACCAAAAAACAACAACA	1740
Db	1681	AATATAAGACAGAAAAATATGTCTTGGTTTTCATTTGTCTTACCAAAAAACAACAACA	1740
QY	1741	AAAAAGTTGTCCTTTGAGAACTTCACCTGCTCCTATGTGGGTACCTGAAGTCAAAATGTC	1800
Db	1741	AAAAAGTTGTCCTTTGAGAACTTCACCTGCTCCTATGTGGGTACCTGAAGTCAAAATGTC	1800
QY	1801	ATTTTGTCTCTGTGAAAAATAAATTTCCCTTCTGTATCCATTTCTGTTAGTTTACTAA	1860
Db	1801	ATTTTGTCTCTGTGAAAAATAAATTTCCCTTCTGTATCCATTTCTGTTAGTTTACTAA	1860
QY	1861	ATCTGTAATACTGTATTTTCTGTTATTCCAAATTTGATGAAGTGAACATCCAATTT	1920
Db	1861	ATCTGTAATACTGTATTTTCTGTTATTCCAAATTTGATGAAGTGAACATCCAATTT	1920
QY	1921	GAAAGTTTGTGTCGACGCTGCTAGCTTAATGAATGTGTTCTATTTGCTTATACATT	1980
Db	1921	GAAAGTTTGTGTCGACGCTGCTAGCTTAATGAATGTGTTCTATTTGCTTATACATT	1980
QY	1981	TATATTAATAAATGTGACATTTTCTAATT 2010	
Db	1981	TATATTAATAAATGTGACATTTTCTAATT 2010	

RESULT 3				
AX697259				
LOCUS	AX697259	2010 bp	DNA	linear
DEFINITION	Sequence 327 from Patent WO0078961.			PAT 02-APR-2003
ACCESSION	AX697259			
VERSION	AX697259.1	GI:29498418		
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			

REFERENCE
AUTHORS
1
Ferrara, N., Stewart, T. A., Williams, P. M., Baker, K. P., Desnoyers, L.,
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

TITLE
Eaton, D. L., Gao, W. Q., Pan, J., Botstein, D., Fong, S., Goddard, A.,
Godowski, P. J., Gurney, A. L., Smith, V., Tumas, D., Wood, W. I.,
Grimaldi, C. J., Hillan, K. J., Paoni, N. F., Roy, M. A. and Watanabe, C. K.
Secreted and transmembrane polypeptides and nucleic acids encoding

JOURNAL Patent: WO 0078961-A 327 28-DEC-2000;
Genentech Inc. (US)

```

FEATURES
    source      location/Qualifiers
               1..2010
               /organism="Homo sapiens"
               /mol_type="unassigned DNA"
               /db_xref="taxon:9606"

```

ORIGIN

Query Match	100.0%	Score 2010;	DB 6;	Length 2010;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 2010; Conservative	0;	Mismatches	0;	Gaps 0;

OY		1	GGA	A	A	A	A	C	T	G	T	C	T	C	T	G	T	G	C	A	C	A	G	A	G	A	C	C	T	G	C	T	C	A	A	G	C	A	G	T	A	G	C	A	G	T	T	C		60
Db		1	GGA	A	A	A	A	C	T	G	T	C	T	C	T	G	T	G	C	A	C	A	G	A	G	A	C	C	T	G	C	T	C	A	A	G	C	A	G	T	A	G	C	A	G	T	T	C		60

QY 61 CGAGTCCAGCTGGCTAAAATCATCCAGAGATATGGCAACCATGCTTAGAATC 120
|||||
Db 61 CGAGTCCAGCTGGCTAAAATCATCCAGAGATATGGCAACCATGCTTAGAATC 120

QY 121 GCTGGGCTGTTCTTG TGATGTGGGCAAGTGGCTGTCACTGCATGCTT 180
|||
Db 121 GCTGGGCTGTTCTTG TGATGTGGGCAAGTGGCTGTCACTGCATGCTT 180

OY	181 CAGTGAAGATGTCGCCCTTCATTGAAACAACATCGTGTTTGGAAA	240
Db	181 CAGTGAAGATGTCGCCCTTCATTGAAACAACATCGTGTTTGGAAA	240

QY	241	GGACTGTGATGAATTGCGTGAGGCAGAGGCTTAACATCAGAGATGCAGTGCAAAATCTATGAT	300
Db	241	GGACTGTGATGAATTGCGTGAGGCAGAGGCTTAACATCAGAGATGCAGTGCAAAATCTATGAT	300
QY	301	TCCCTGCTGGCTCTTTCTCCGGACCTACAGGCAGCCAGAGACTGATGTGTGCTTCC	360
Db	301	TCCCTGCTGGCTCTTTCTCCGGACCTACAGGCAGCCAGAGACTGATGTGTGCTTCC	360
QY	361	GTGATGTCTTCTTGGCTTTCATGATGGCCATCCTTGGCATGAATGCACCAGGTGCAG	420
Db	361	GTGATGTCTTCTTGGCTTTCATGATGGCCATCCTTGGCATGAATGCACCAGGTGCAG	420
QY	421	GGGGCAATGAGAAGGTGAAGGCTCACATCTGTGACGGCTGGAATCATCTTCATCATC	480
Db	421	GGGGCAATGAGAAGGTGAAGGCTCACATCTGTGACGGCTGGAATCATCTTCATCATC	480
QY	481	ACGGGCATGTGTGTCTCATCCCTGTGAGCTGGGTGCCAATGCCATCATCAGAGATTTC	540
Db	481	ACGGGCATGTGTGTCTCATCCCTGTGAGCTGGGTGCCAATGCCATCATCAGAGATTTC	540
QY	541	TATACTCAATAGTGAATGTTGCCCAAAAAGTGAGCTTGGAGAAGCTCTCTACTTGA	600
Db	541	TATACTCAATAGTGAATGTTGCCCAAAAAGTGAGCTTGGAGAAGCTCTCTACTTGA	600
QY	601	TGACCAACGGCACTGTGTCTGATTTGTGGAGAGCTCTGTCTGTCTGCGTTTGTGTC	660
Db	601	TGACCAACGGCACTGTGTCTGATTTGTGGAGAGCTCTGTCTGTCTGCGTTTGTGTC	660
QY	661	AACGAAAAGACGATGATACAGATACTCGATACCTTCCCATGCACAACCCAAAAAGT	720
Db	661	AACGAAAAGACGATGATACAGATACTCGATACCTTCCCATGCACAACCCAAAAAGT	720
QY	721	TATCACACCGAAAAGATCACCGAGCGTCTACTCCAGAAGTCATGTGTAGTTGTGT	780
Db	721	TATCACACCGAAAAGATCACCGAGCGTCTACTCCAGAAGTCATGTGTAGTTGTGT	780
QY	781	ATGTTTTTTTAACTTTACTATAAAGCCATGCAAAATGACAAAAATCTATTTACTTTC	840
Db	781	ATGTTTTTTTAACTTTACTATAAAGCCATGCAAAATGACAAAAATCTATTTACTTTC	840
QY	841	AAATGACCCCAAGAACTTTGATTTACTGTCTTAACTGCTTAATCTTAATTACAGA	900
Db	841	AAATGACCCCAAGAACTTTGATTTACTGTCTTAACTGCTTAATCTTAATTACAGA	900
QY	901	ACTGTGATCAGCTATTATGATTTCTATAAGCTATTTGACGAGATGAGATTTAAACC	960
Db	901	ACTGTGATCAGCTATTATGATTTCTATAAGCTATTTGACGAGATGAGATTTAAACC	960
QY	961	AATGCTTTGATTTGTTCTAGAAAGTATAGTAATTTGTTTCTAAGGTGTTCAAGCATCTA	1020
Db	961	AATGCTTTGATTTGTTCTAGAAAGTATAGTAATTTGTTTCTAAGGTGTTCAAGCATCTA	1020
QY	1021	CTCTTTTATCATTTACTTCAAAATGACATTGCTAAAGACTGCATATTTTACTACTGA	1080
Db	1021	CTCTTTTATCATTTACTTCAAAATGACATTGCTAAAGACTGCATATTTTACTACTGA	1080
QY	1081	ATTTCTCCACGACATAGCATTTATGTACATAGATGAGTGAACATTTATCTCACATAGA	1140
Db	1081	ATTTCTCCACGACATAGCATTTATGTACATAGATGAGTGAACATTTATCTCACATAGA	1140
QY	1141	GACATGCTTATATGTTTAAATGAATGCCAGTCCATTACACTGAATAAATAGA	1200
Db	1141	GACATGCTTATATGTTTAAATGAATGCCAGTCCATTACACTGAATAAATAGA	1200
QY	1201	ACTCAACTATTGCTTTTCAGGGAATCATGGAATAGGTTGAAGAAGTTACTATTAATTG	1260
Db	1201	ACTCAACTATTGCTTTTCAGGGAATCATGGAATAGGTTGAAGAAGTTACTATTAATTG	1260
QY	1261	TTTAAAAACAGCTTAGGATTAATGTCTCCATTTTATATGAAGATTTAAATGAAGCTT	1320
Db	1261	TTTAAAAACAGCTTAGGATTAATGTCTCCATTTTATATGAAGATTTAAATGAAGCTT	1320

QY	1321	TAATCAGCATTTGTAAGAAATTTGAATGGCTTCTGATATGCTGTTTTTTAGCCTAGAG	1380
Db	1321	TAATCAGCATTTGTAAGAAATTTGAATGGCTTCTGATATGCTGTTTTTTAGCCTAGAG	1380
QY	1381	TTAGAAATCCTAACTCTTTATTCCTCTCTCCAGAGGCTTTTTTTCTGTGTATTA	1440
Db	1381	TTAGAAATCCTAACTCTTTATTCCTCTCTCCAGAGGCTTTTTTTCTGTGTATTA	1440
QY	1441	ATTACATTTTAAACCGAGATATTTGTCAAGGGCTTTGCATTCAACCTGCTTTCC	1500
Db	1441	ATTACATTTTAAACCGAGATATTTGTCAAGGGCTTTGCATTCAACCTGCTTTCC	1500
QY	1501	AGGGCTATCTCAGAGAAAGATAAAGTGTGATCTAAGAAAAAGTGATGCTTTAGGA	1560
Db	1501	AGGGCTATCTCAGAGAAAGATAAAGTGTGATCTAAGAAAAAGTGATGCTTTAGGA	1560
QY	1561	AGTGAATAATTTTTGTTTTTGTATTTGAAGAAGATGATGCAATTTTGACAAGAAATCAT	1620
Db	1561	AGTGAATAATTTTTGTTTTTGTATTTGAAGAAGATGATGCAATTTTGACAAGAAATCAT	1620
QY	1621	ATATGTATGATATATTTTAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATATA	1680
Db	1621	ATATGTATGATATATTTTAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATATA	1680
QY	1681	AATAAAAAGACGAGAAAAATATGCTCTGTTTCAATTTGCTTACCAAAAAACAACAA	1740
Db	1681	AATAAAAAGACGAGAAAAATATGCTCTGTTTCAATTTGCTTACCAAAAAACAACAA	1740
QY	1741	AAAAAGTTGTCCTTTGGAACCTTCACTGCTCTATGTGGTACCTGAGTCAAAATTGTC	1800
Db	1741	AAAAAGTTGTCCTTTGGAACCTTCACTGCTCTATGTGGTACCTGAGTCAAAATTGTC	1800
QY	1801	ATTTTGTCTGTGAAAAATAAATTTCTCTTGTACCAATTTCTGTTAGTTTACTAAA	1860
Db	1801	ATTTTGTCTGTGAAAAATAAATTTCTCTTGTACCAATTTCTGTTAGTTTACTAAA	1860
QY	1861	ATCTGTAATACTGTATTTTCTGTTTATTCCAAATTTGATGAACCTGACAAATCCAATT	1920
Db	1861	ATCTGTAATACTGTATTTTCTGTTTATTCCAAATTTGATGAACCTGACAAATCCAATT	1920
QY	1921	GAAAGTTTGTGTGACGCTGTCTAGCTTAAATGAATGTCTTATTTGCTTTATACATT	1980
Db	1921	GAAAGTTTGTGTGACGCTGTCTAGCTTAAATGAATGTCTTATTTGCTTTATACATT	1980
QY	1981	TATATTATAAATTGTACATTTTCTAATT	2010
Db	1981	TATATTATAAATTGTACATTTTCTAATT	2010

RESULT 4
AY358707 2010 bp mRNA linear PRI 03-OCT-2003
LOCUS Homo sapiens clone DNA73735 CLDN8 (UNQ779) mRNA, complete cds.
DEFINITION
ACCESSION AY358707
VERSION AY358707.1 GI:37182535
KEYWORDS FLI CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2010)
AUTHORS Clark,H.F., Gurney,A.L., Abaya,E., Baker,K., Baldwin,D., Brush,J., Chen,J., Chow,B., Chui,C., Crowley,C., Currell,B., Deuel,B., Dowd,P., Eaton,D., Foster,J., Grimaldi,C., Gu,Q., Hass,P.E., Heldens,S., Huang,A., Kim,H.S., Klimowski,L., Jin,Y., Johnson,S., Lee,J., Lewis,L., Liao,D., Mark,M., Robble,E., Sanchez,C., Schoenfeld,J., Seshagiri,S., Simmons,L., Singh,J., Smith,V., Stinson,J., Vagts,A., Vandlen,R., Watanabe,C., Wieand,D., Woods,K., Xie,M.H., Yansura,D., Yi,S., Yu,G., Yuan,J., Zhang,M., Zhang,Z., Goddard,A., Wood,W.I. and Godowski,P.
TITLE The Secreted Protein Discovery Initiative (SPDI), a Large-Scale Effort to Identify Novel Human Secreted and Transmembrane Proteins: A Bioinformatics Assessment

JOURNAL	Genome Res. 13 (10), 2265-2270 (2003)	
PUBMED	12975309	
REFERENCE	2 (bases 1 to 2010)	
AUTHORS	Clark,H.F.	
TITLE	Direct Submission	
JOURNAL	Submitted (01-AUG-2003) Department of Bioinformatics, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA	
FEATURES	Location/Qualifiers	
source	1..2010 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="DNA73735" 1..2010 /locus_tag="UNQ779" 97..774 /locus_tag="UNQ779" /note="PRO1573" /codon_start=1 /product="CLDN8" /protein_id="AAQ89070.1" /db_xref="GI:37182536" /translation="MATHALEIAGLFLGVGVGTVAVTMPQWRVSAFIENNIIVFEE NFEGLMNCVRQANIRMQCKIYDSLALSPDLQARGLMCAASVMSFLAFMVAIILGM KCTRCTGDNEKVKHILLTAGIIFITGMVLIIPSWVANAIIRDFTNSIVVAQKRE LGEALYLGMWTLVLIIVGALFCCVFCNFKSSSYRYSISPSHRTQKSYHTGKSPSV YRSQYV"	
gene		
CDS		
Query Match	100.0%; Score 2010; DB 9; Length 2010;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 2010; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 GGAATAACTGTTCTCTCTGTGGGACACAGAAACCTGCTTCAAGCAGAAAGTAGCAGTTCC	60
Db	1 GGAATAACTGTTCTCTCTGTGGGACACAGAAACCTGCTTCAAGCAGAAAGTAGCAGTTCC	60
QY	61 CGAGTCCAGCTGGCTAAACTCATCCAGAGGATTAATGGCAACCCATGCCCTTAGAAATC	120
Db	61 CGAGTCCAGCTGGCTAAACTCATCCAGAGGATTAATGGCAACCCATGCCCTTAGAAATC	120
QY	121 GCTGGGCTGTTCTTGTGTGTGTGAATGTGGGACACAGTGGCTGTCACTGTCAATGCTT	180
Db	121 GCTGGGCTGTTCTTGTGTGTGTGAATGTGGGACACAGTGGCTGTCACTGTCAATGCTT	180
QY	181 CAGTGAAGTGTGCGCCCTTCATTTGAACAACATCGTGTGTTTTGAAAACTTCTGGGAA	240
Db	181 CAGTGAAGTGTGCGCCCTTCATTTGAACAACATCGTGTGTTTTGAAAACTTCTGGGAA	240
QY	241 GCACTGTGATGAATTTGCGTGAAGGACGCTAACATCAGAGTCAAGTCAAAATCTATGAT	300
Db	241 GCACTGTGATGAATTTGCGTGAAGGACGCTAACATCAGAGTCAAGTCAAAATCTATGAT	300
QY	301 TCCCTGCTGCTCTTTCTCCGACCTACAGGACGACAGAGACTGATGTGTGCTGCTTCC	360
Db	301 TCCCTGCTGCTCTTTCTCCGACCTACAGGACGACAGAGACTGATGTGTGCTGCTTCC	360
QY	361 GTGATGCTCTTCTTGCTTCATGATGGCCATCTTGCGATGAATGCAACGAGTGCACG	420
Db	361 GTGATGCTCTTCTTGCTTCATGATGGCCATCTTGCGATGAATGCAACGAGTGCACG	420
QY	421 GGGACAATGAGAAGGTGAAGGCTCACATTTGCTGACGGCTGGAATCATCTTCATCATC	480
Db	421 GGGACAATGAGAAGGTGAAGGCTCACATTTGCTGACGGCTGGAATCATCTTCATCATC	480
QY	481 ACGGCGATGCTGTGCTCATCTCTGTGAGCTGGGTTGCCAATGCCATCATCAGAGATTTC	540
Db	481 ACGGCGATGCTGTGCTCATCTCTGTGAGCTGGGTTGCCAATGCCATCATCAGAGATTTC	540
QY	541 TATTACTCAATAAGTGAATGTTGCCCAAAACGTGACCTTGGAAGAGCTCTACTTAGGA	600
Db	541 TATTACTCAATAAGTGAATGTTGCCCAAAACGTGACCTTGGAAGAGCTCTACTTAGGA	600

QY 601 TGAACCAACGCGACTGCTGCTGATGTTGGAGAGCTCTGTTCTGCTGCGTTTGTGTC 660
DB 601 TGAACCAACGCGACTGCTGCTGATGTTGGAGAGCTCTGTTCTGCTGCGTTTGTGTC 660
QY 661 AACGAAAAGACAGTAGCTACAGATACCTGATCCATCGCACCAACCAAAAAAGT 720
DB 661 AACGAAAAGACAGTAGCTACAGATACCTGATCCATCGCACCAACCAAAAAAGT 720
QY 721 TATCACACCGGAAAGAGTCACCGAGCGTCTACTCCAGAGTCAGTATGTGTGTGT 780
DB 721 TATCACACCGGAAAGAGTCACCGAGCGTCTACTCCAGAGTCAGTATGTGTGTGT 780
QY 781 ATGTTTTTTTAACTTTACTATAAAGCCATGCAAAATGACAAAATCTATATTACTTCTCA 840
DB 781 ATGTTTTTTTAACTTTACTATAAAGCCATGCAAAATGACAAAATCTATATTACTTCTCA 840
QY 841 AAATGGAACCCCAAGAACTTTGATTTACTGTTCTTAAGTCCCTAATCTTAATTAACAGA 900
DB 841 AAATGGAACCCCAAGAACTTTGATTTACTGTTCTTAAGTCCCTAATCTTAATTAACAGA 900
QY 901 ACTGTCATCAGCTATTATGATTTCTATAAGCTATTTCAGCAGAAATGAGATATTAAACCC 960
DB 901 ACTGTCATCAGCTATTATGATTTCTATAAGCTATTTCAGCAGAAATGAGATATTAAACCC 960
QY 961 AATGCTTTGATGTTCTAGAAAGTATAGTAATTTGTTTCTAAGGTGTTCAAGCATCTA 1020
DB 961 AATGCTTTGATGTTCTAGAAAGTATAGTAATTTGTTTCTAAGGTGTTCAAGCATCTA 1020
QY 1021 CTCTTTTATCATTTACTTCAAAATGACATTGCTAAAGACTGCATTATTTTACTACTGTA 1080
DB 1021 CTCTTTTATCATTTACTTCAAAATGACATTGCTAAAGACTGCATTATTTTACTACTGTA 1080
QY 1081 ATTTCTCCACGACATAGCATTTATGATAGATGAGTGTAAACATTTATATCTCACATAGA 1140
DB 1081 ATTTCTCCACGACATAGCATTTATGATAGATGAGTGTAAACATTTATATCTCACATAGA 1140
QY 1141 GACATGCTTATATGGTTTATTAAATGAAATGCCAGTCCATTACACTGAAATTAATAGA 1200
DB 1141 GACATGCTTATATGGTTTATTAAATGAAATGCCAGTCCATTACACTGAAATTAATAGA 1200
QY 1201 ACTCAACTATTGCTTTTCAGGGAATCATGAGTAGGGTTGAAGAAAGTTACTATTAAATG 1260
DB 1201 ACTCAACTATTGCTTTTCAGGGAATCATGAGTAGGGTTGAAGAAAGTTACTATTAAATG 1260
QY 1261 TTTAAAAAACAGCTTAGGATTAAATGCTCTCCATTATTAATGAAGATTAAATGAAGGCTT 1320
DB 1261 TTTAAAAAACAGCTTAGGATTAAATGCTCTCCATTATTAATGAAGATTAAATGAAGGCTT 1320
QY 1321 TAAATCAGCATTTGTAAGAAATGAATGGCTTCTGATATGCTGTTTTTACGCTAGAG 1380
DB 1321 TAAATCAGCATTTGTAAGAAATGAATGGCTTCTGATATGCTGTTTTTACGCTAGAG 1380
QY 1381 TTAGAATCCTTAACCTCTTATCCTCTCTCCAGAGGCTTTTCTTGTGATTTAA 1440
DB 1381 TTAGAATCCTTAACCTCTTATCCTCTCTCCAGAGGCTTTTCTTGTGATTTAA 1440
QY 1441 ATTAACATTTTAAACGAGATATTTTGTCAAGGGCTTTCATTCAAACTGCTTTCC 1500
DB 1441 ATTAACATTTTAAACGAGATATTTTGTCAAGGGCTTTCATTCAAACTGCTTTCC 1500
QY 1501 AGGGCTATACTCAGAGAAGATATAAAGTGTGATCTAAGAAAAAGTGATGTTTAGGAA 1560
DB 1501 AGGGCTATACTCAGAGAAGATATAAAGTGTGATCTAAGAAAAAGTGATGTTTAGGAA 1560
QY 1561 AGTGAATAATTTTGTGTTTGTATTTGAAGAAGATGATGATTTTGACAGAATAATCAT 1620
DB 1561 AGTGAATAATTTTGTGTTTGTATTTGAAGAAGATGATGATTTTGACAGAATAATCAT 1620
QY 1621 AATATGATGATATATTTTAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATATA 1680
DB 1621 AATATGATGATATATTTTAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATATA 1680
QY 1681 AATATAAAGACGAGAAAAATATGCTTGTTTCAATTGCTTACCAAAAAACAACAACA 1740

DB 1681 AATATAAAGACGAGAAAAATATGCTTGTTTCAATTGCTTACCAAAAAACAACAACA 1740
QY 1741 AAAAGTGTCTTTGAGAACTTCACTGCTCTATGTGGTAACTGAGTCAAAATTGTC 1800
DB 1741 AAAAGTGTCTTTGAGAACTTCACTGCTCTATGTGGTAACTGAGTCAAAATTGTC 1800
QY 1801 ATTTTGTCTGTGAATAATAATTCTCTCTGTGTACCATTTCTGTTAGTTTAACTAAA 1860
DB 1801 ATTTTGTCTGTGAATAATAATTCTCTCTGTGTACCATTTCTGTTAGTTTAACTAAA 1860
QY 1861 ATCTGTAATACTGTATTTTCTGTTTATCCAAATTGATGAACCTGACAAATCAATT 1920
DB 1861 ATCTGTAATACTGTATTTTCTGTTTATCCAAATTGATGAACCTGACAAATCAATT 1920
QY 1921 GAAAGTTGTGTCGACGCTGTCTAGCTTAAATGAATGTGTTCTATTGCTTTATACATT 1980
DB 1921 GAAAGTTGTGTCGACGCTGTCTAGCTTAAATGAATGTGTTCTATTGCTTTATACATT 1980
QY 1981 TATATTAATAATTTGTACATTTTCTAATT 2010
DB 1981 TATATTAATAATTTGTACATTTTCTAATT 2010

RESULT 5
AP000884/c 137955 bp DNA linear PRI 16-DEC-1999
LOCUS AP000884 Homo sapiens genomic DNA, chromosome 21q22.1, clone: B463J19,
DEFINITION SOD-AML region, complete sequence.
ACCESSION AP000884
VERSION AP000884.1 GI:6580113
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE Published Only in Database (1999)
JOURNAL 2 (bases 1 to 137955)
REFERENCE Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
AUTHORS

TITLE Direct Submission
JOURNAL Submitted (15-DEC-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagami-hara, Kanagawa 228-8555,
Japan (E-mail: hattori@gs.c.riken.go.jp,
URL: http://hgp.gs.c.riken.go.jp/, Tel: 81-42-778-9923,
Fax: 81-42-778-9924)

FEATURES
source location/Qualifiers
1..137955
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="B463J19"

ORIGIN

Query Match 99.9%; Score 2008.4; DB 9; Length 137955;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2009; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GGAATACTGTTCTCTTCTGTGGCAGAGAACCCCTGCTTCAAGCAGAGTAGAGTTC 60
DB 58357 GGAATACTGTTCTCTTCTGTGGCAGAGAACCCCTGCTTCAAGCAGAGTAGAGTTC 58298
QY 61 CGAGTTCAGCTGGCTAAACTCATCCGAGAGATATGGCAACCCATGCCTTAGAATC 120
DB 58297 CGAGTTCAGCTGGCTAAACTCATCCGAGAGATATGGCAACCCATGCCTTAGAATC 58238

QY	121	GCTGGGCTGTTTCTTGGTGGTGTGGATGGTGGGCACAGTGGCTGTCACTGTCAATGCT	180
Db	58237	GCTGGGCTGTTTCTTGGTGGTGTGGATGGTGGGCACAGTGGCTGTCACTGTCAATGCT	58178
QY	181	CAGTGGAGAGTGTGGCTTCAATGAAAAACAATGTGTTTGGAAAACTTCTGGAA	240
Db	58177	CAGTGGAGAGTGTGGCTTCAATGAAAAACAATGTGTTTGGAAAACTTCTGGAA	58118
QY	241	GGACTGTGATGAATGCGTGAGGCAGGCTAACATCAGATGCAATGCAAAATCTATGAT	300
Db	58117	GGACTGTGATGAATGCGTGAGGCAGGCTAACATCAGATGCAATGCAAAATCTATGAT	58058
QY	301	TCCCTGCTGGCTCTTCTCCGACCTACAGCAGCCAGAGACTGATGTGTGCTGTCC	360
Db	58057	TCCCTGCTGGCTCTTCTCCGACCTACAGCAGCCAGAGACTGATGTGTGCTGTCC	57998
QY	361	GTGATGTCTTCTTGAGCTTTCATGATGGCCATCCTTGGCATGAATGCCACAGGTGACG	420
Db	57997	GTGATGTCTTCTTGAGCTTTCATGATGGCCATCCTTGGCATGAATGCCACAGGTGACG	57938
QY	421	GGGACAAATGAGAAGTGAAGCTCACATTCGTGACGGCTGAAATCATCTTCATCATC	480
Db	57937	GGGACAAATGAGAAGTGAAGCTCACATTCGTGACGGCTGAAATCATCTTCATCATC	57878
QY	481	ACGGGCATGGTGTGCTCATCCCTGTGAGCTGGGTTGCCAATGCCATCATCAGATTTTC	540
Db	57877	ACGGGCATGGTGTGCTCATCCCTGTGAGCTGGGTTGCCAATGCCATCATCAGATTTTC	57818
QY	541	TATACTCAATAGTGAATGTTGCCCAAAAAGTGAAGCTTGGAGAAAGCTCTACTTATGA	600
Db	57817	TATACTCAATAGTGAATGTTGCCCAAAAAGTGAAGCTTGGAGAAAGCTCTACTTATGA	57758
QY	601	TGGACCAAGGCACTGGTGTGATGTTGGAGAGAGCTGTGTTCTGCTGCGTTTGTGTC	660
Db	57757	TGGACCAAGGCACTGGTGTGATGTTGGAGAGAGCTGTGTTCTGCTGCGTTTGTGTC	57698
QY	661	AACGAAAAGAGCAGTAGCTACAGATACGTATACCTTCCCATCGACAACCCAAAAAGT	720
Db	57697	AACGAAAAGAGCAGTAGCTACAGATACGTATACCTTCCCATCGACAACCCAAAAAGT	57638
QY	721	TATCACACCGGAAAGAGTCAACCGAGCGTCTACTCCAGAGTCAATATGTAGTTGTGT	780
Db	57637	TATCACACCGGAAAGAGTCAACCGAGCGTCTACTCCAGAGTCAATATGTAGTTGTGT	57578
QY	781	ATGTTTTTTTAACTTTACTATTAAGCCATGCAAAATGACAAAACTATATTACTTTCTCA	840
Db	57577	ATGTTTTTTTAACTTTACTATTAAGCCATGCAAAATGACAAAACTATATTACTTTCTCA	57518
QY	841	AAATGACCCCAAGAAACTTTGATTTACTGTTCTTAACCTGCTAATCTTAATTACAGA	900
Db	57517	AAATGACCCCAAGAAACTTTGATTTACTGTTCTTAACCTGCTAATCTTAATTACAGA	57458
QY	901	ACTGTGCATCAGCTATTATGATTTCTATAAGCTATTTTCAGCAGAAATGAGATTAAACC	960
Db	57457	ACTGTGCATCAGCTATTATGATTTCTATAAGCTATTTTCAGCAGAAATGAGATTAAACC	57398
QY	961	AATGCTTTGATGTTTCTAGAAAGTATAGTAATTTGTTTCTAAGGTGTTCAAGCATCTA	1020
Db	57397	AATGCTTTGATGTTTCTAGAAAGTATAGTAATTTGTTTCTAAGGTGTTCAAGCATCTA	57338
QY	1021	CTCTTTTATCATTTTACTTCAAAATGACATTTGCTAAAGCTGCATTAATTTTACTACTGTA	1080
Db	57337	CTCTTTTATCATTTTACTTCAAAATGACATTTGCTAAAGCTGCATTAATTTTACTACTGTA	57278
QY	1081	ATTCTCTCAGCAGATAGCATATGTACATAGATGAGTGTAACTTATATCTCACATAGA	1140
Db	57277	ATTCTCTCAGCAGATAGCATATGTACATAGATGAGTGTAACTTATATCTCACATAGA	57218
QY	1141	GACATGCTTATATGTTTTTATTAAATGAATGCCAGTCCATTTACCTGAATAAATAGA	1200
Db	57217	GACATGCTTATATGTTTTTATTAAATGAATGCCAGTCCATTTACCTGAATAAATAGA	57158
QY	1201	ACTCAACTATTTGCTTTTACGGGAATCATGATAGGGTTGAAGAGGTTACTATTAAATG	1260

Db	57157	ACTCACTATTCGCTTTTCAGGGAATCATGAGTAGGCTTGAGAAAGTTACTATTAATTG	57098
QY	1261	TTTAAAAACAGCTTAGGATTAATGTCCTCCATTTAATGAAGATTAAATGAAGCTT	1320
Db	57097	TTTAAAAACAGCTTAGGATTAATGTCCTCCATTTAATGAAGATTAAATGAAGCTT	57038
QY	1321	TAATCAGCATTGTAAAGGAAATGAAATGGCTTTCGATATGCTGTTTTTAGCCTAGAG	1380
Db	57037	TAATCAGCATTGTAAAGGAAATGAAATGGCTTTCGATATGCTGTTTTTAGCCTAGAG	56978
QY	1381	TTAGAAATCCTAACTTCTTATCCTCTCTCCAGAGGCTTTTTTCTGTGTAATA	1440
Db	56977	TTAGAAATCCTAACTTCTTATCCTCTCTCCAGAGGCTTTTTTCTGTGTAATA	56918
QY	1441	ATTAACATTTTAAAAACGAGATATTTGTCAAGGGCTTGCATTCAAACTGCTTTCC	1500
Db	56917	ATTAACATTTTAAAAACGAGATATTTGTCAAGGGCTTGCATTCAAACTGCTTTCC	56858
QY	1501	AGGCTATACTCAGAGAAAGATAAAGTGTGATCTAAGAAAAAGTGATGTTTTAGGAA	1560
Db	56857	AGGCTATACTCAGAGAAAGATAAAGTGTGATCTAAGAAAAAGTGATGTTTTAGGAA	56798
QY	1561	AGTGAAATATTTTTGTTTTGTATTTGAAGAAGATGATGCAATTTGACAGAATCAT	1620
Db	56797	AGTGAAATATTTTTGTTTTGTATTTGAAGAAGATGATGCAATTTGACAGAATCAT	56738
QY	1621	ATATGTATGATATATTTTAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATATA	1680
Db	56737	ATATGTATGATATATTTTAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATATA	56678
QY	1681	AATAAAAAGCAGAAAAAATATGTCCTGTTTTCATTTGCTTACCAAAAAACAACAACAA	1740
Db	56677	AATAAAAAGCAGAAAAAATATGTCCTGTTTTCATTTGCTTACCAAAAAACAACAACAA	56618
QY	1741	AAAAAGTTGTCCTTGAGAACTTCACTGCTCCTATGTGGTACCTGAGTCAAAATTGTC	1800
Db	56617	AAAAAGTTGTCCTTGAGAACTTCACTGCTCCTATGTGGTACCTGAGTCAAAATTGTC	56558
QY	1801	ATTTTGTCTGTGAAAAAATAAATTCCTTCTTGTAACAATTCTGTTAGTTTTACTAAA	1860
Db	56557	ATTTTGTCTGTGAAAAAATAAATTCCTTCTTGTAACAATTCTGTTAGTTTTACTAAA	56498
QY	1861	ATCTGTAATACTGTATTTTCTGTTTATTCCAAATTTGATGAAGAACTGACAATCCAATT	1920
Db	56497	ATCTGTAATACTGTATTTTCTGTTTATTCCAAATTTGATGAAGAACTGACAATCCAATT	56438
QY	1921	GAAAGTTTGTGTCGACGCTGTGCTTACCTTAATGAATGTGTTCTATTGCTTATACATT	1980
Db	56437	GAAAGTTTGTGTCGACGCTGTGCTTACCTTAATGAATGTGTTCTATTGCTTATACATT	56378
QY	1981	TATATTAAATAATTGTACATTTTCTAATT 2010	
Db	56377	TATATTAAATAATTGTACATTTTCTAATT 56348	

RESULT 6	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
AP001707/c	AP001707	Homo sapiens genomic DNA, chromosome 21q, section 51/105.	AP001707 AL163252 BA000005	AP001707.1	GI:7768786	Homo sapiens (human)	Homo sapiens	1	Hattori, M., Fujiyama, A., Taylor, T. D., Watanabe, H., Yada, T., Park, H. S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D. K., Soeda, E., Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K., Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R., Patterson, D., Reichwald, K., Rump, A., Schilhabel, M., Schudy, A.,

Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K., Asakawa,S., Shintani,A., Sasaki,K., Nagamine,K., Mitsuayama,S., Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G., Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S., Hennig,S., Rieseemann,L., Dagand,E., Wehmeyer,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Yaspo,M.L.
The DNA sequence of human chromosome 21
Nature 405 (6784), 311-319 (2000)
20289799
PUBMED
10830953
2 (bases 1 to 340000)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Reichwald,K., Rump,A., Schilhabel,M., Schudy,A., Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K., Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuayama,S., Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G., Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S., Hennig,S., Rieseemann,L., Dagand,E., Wehmeyer,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Yaspo,M.L.
Direct Submission
Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)
On Jan 16, 2002 this sequence version replaced gi:7717317.
The chromosome 21 mapping and sequencing consortium consisting of * RIKEN Genomic Sciences Center, Human Genome Research Group, * Sagamihara 228-8555, Japan,
* e.mail: hattori@gsc.riken.go.jp/
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Molecular Biology, * Tokyo 160-8582, Japan,
* e.mail: nshimizu@mb-med.keio.ac.jp
* URL: http://www.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail: info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
* Ihmestrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/
AL163252: Submitted (10-APR-2000).
Location/Qualifiers
1..340000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
<1..148
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="R355D16, 5' partial"

/clone_lib="RPCI-11 BAC library"
/note="Accession No. AP000236"
<1..48622
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="R347N6, 5' partial"
/clone_lib="RPCI-11 BAC library"
/note="Accession No. AP000390"
<1..57177
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="p681K20, 5' partial"
/clone_lib="RPC11,3-5 PAC library"
/note="Accession No. AF191069"
26572..135558
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="p131P23"
/clone_lib="RPC11,3-5 PAC library"
/note="Accession No. AF190641"
48469..158332
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="p209K9"
/clone_lib="RPC11,3-5 PAC library"
/note="Accession No. AF096876"
156690..294644
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="R463J19"
/clone_lib="RPCI-11 BAC library"
/note="Accession No. AP000884"
292534..>340000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="R680H4, 3' partial"
/clone_lib="RPCI-11 BAC library"
/note="Accession No. AP000885"
1610..1766
/note="MER50"
/rpt_family="LTR/MER4-group"
/rpt_type="DISPERSED
1857..2051
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type="DISPERSED
2054..2392
/note="MER50"
/rpt_family="LTR/MER4-group"
/rpt_type="DISPERSED
3327..3518
/note="MER102"
/rpt_family="DNA/MER1_type"
/rpt_type="DISPERSED
complement(3582..3694)

repeat_region	/note="L2"									
	/rpt_family="LINE/L2"									
	/rpt_type=DISPERSED									
	4580. .4614									
repeat_region	/note="(TTC)n"									
	/rpt_family="Simple_repeat"									
	/rpt_type=TANDEM									
	5743. .6052									
repeat_region	/note="AluSq"									
	/rpt_family="SINE/Alu"									
	/rpt_type=DISPERSED									
	complement(6134. .6214)									
repeat_region	/note="MIR"									
	/rpt_family="SINE/MIR"									
	/rpt_type=DISPERSED									
	complement(6299. .6651)									
repeat_region	/note="MLT1A1"									
	/rpt_family="LTR/MaLR"									
	/rpt_type=DISPERSED									
	7531. .7839									
repeat_region	/note="AluSc"									
	/rpt_family="SINE/Alu"									
	/rpt_type=DISPERSED									
	9145. .9173									
repeat_region	/note="(A)n"									
	/rpt_family="Simple_repeat"									
	/rpt_type=TANDEM									
	complement(10285. .10467)									
repeat_region	/note="MIR"									
	/rpt_family="SINE/MIR"									
	/rpt_type=DISPERSED									
	10468. .10753									
repeat_region	/note="AluJb"									
	/rpt_family="SINE/Alu"									
	/rpt_type=DISPERSED									
	10757. .10785									
repeat_region	/note="(TAAAA)n"									
	/rpt_family="Simple_repeat"									
	/rpt_type=TANDEM									
	complement(10785. .10800)									
Query Match 99.9%; Score 2008.4; DB 9; Length 340000;										
Best Local Similarity 100.0%; Pred. No. 0;										
Matches 2009; Conservative 0; Mismatches 1; Indels 0; Gaps 0;										
QY	1	GGAAAACTGTTCTCTTCTGTGGCACAGAGAACCTTGCTTCAAAGCAGAGAGTAGCAGTTC	60							
Db	215046	GGAAAACTGTTCTCTTCTGTGGCACAGAGAACCTTGCTTCAAAGCAGAGAGTAGCAGTTC	214987							
QY	61	CGAGTCCAGCTGGCTAAAACTCATCCCAGAGGATAATGGCAACCCATGCCCTTAGAAATC	120							
Db	214986	CGAGTCCAGCTGGCTAAAACTCATCCCAGAGGATAATGGCAACCCATGCCCTTAGAAATC	214927							
QY	121	GCTGGGCTGTTTCTTGTGGTGTGTTGGAATGGTGGGCACAGTGGCTGTCACTGTCAATGCCT	180							
Db	214926	GCTGGGCTGTTTCTTGTGGTGTGTTGGAATGGTGGGCACAGTGGCTGTCACTGTCAATGCCT	214867							
QY	181	CAGTGGAGAGTGTGGCCTTCATTGAAAAACAACATCGTGGTTTTTGAAAAACTTCTGGGAA	240							
Db	214866	CAGTGGAGAGTGTGGCCTTCATTGAAAAACAACATCGTGGTTTTTGAAAAACTTCTGGGAA	214807							
QY	241	GGACTGTGGATGAATTGCGTGAGGCAGGCTAACATCAGGATGCAGTGCAGAAATCTATGAT	300							
Db	214806	GGACTGTGGATGAATTGCGTGAGGCAGGCTAACATCAGGATGCAGTGCAGAAATCTATGAT	214747							
QY	301	TCCCTGTGGCTCTTTTCTCCGGACCTACAGGCAGCCAGAGGACTGATGTGTGCTGCTTCC	360							
Db	214746	TCCCTGTGGCTCTTTTCTCCGGACCTACAGGCAGCCAGAGGACTGATGTGTGCTGCTTCC	214687							
QY	361	GTGATGTCCTTCTTGGCTTTTCATGATGGCCATCCTTGGCATGAAATGCACCGGTGCACG	420							
Db	214686	GTGATGTCCTTCTTGGCTTTTCATGATGGCCATCCTTGGCATGAAATGCACCGGTGCACG	214627							
QY	421	GGGGACAATGAGAAGGTGAAGGCTCACATTCCTGCTGACGGCTGGAATCATCTTCATCATC	480							

Db	214626	GGGACAATGAGAAGGTGAAGGCTCACATTCCTGCTGACGGCTGGAATCATCTTCATCATC	214567							
QY	481	ACGGGATGGTGGTCTCATCCCTGTGAGCTGGGTGCCAATGCCATCATCAGAGATTTC	540							
Db	214566	ACGGGATGGTGGTCTCATCCCTGTGAGCTGGGTGCCAATGCCATCATCAGAGATTTC	214507							
QY	541	TATAACTCAATAGTGAATGTGCCCCAAAAACGTGAGCTTGGAGAAGCTCTCTACTTAGGA	600							
Db	214506	TATAACTCAATAGTGAATGTGCCCCAAAAACGTGAGCTTGGAGAAGCTCTCTACTTAGGA	214447							
QY	601	TGGACCAACGGCACTGGTGTGCTGATTTGTTGGAGGAGCTCTGTTCTGCTGCGTTTTTTTGTTC	660							
Db	214446	TGGACCAACGGCACTGGTGTGCTGATTTGTTGGAGGAGCTCTGTTCTGCTGCGTTTTTTTGTTC	214387							
QY	661	AACGAAAAGACAGTAGTACTACAGATACCTCGATACCTTCCCATCGCACAAACCCAAAAAGT	720							
Db	214386	AACGAAAAGACAGTAGTACTACAGATACCTCGATACCTTCCCATCGCACAAACCCAAAAAGT	214327							
QY	721	TATCACACCGGAAAGAGTCAACGAGCGTCTACTCCAGAAAGTCAGTATGTGTAGTTGTGT	780							
Db	214326	TATCACACCGGAAAGAGTCAACGAGCGTCTACTCCAGAAAGTCAGTATGTGTAGTTGTGT	214267							
QY	781	ATGTTTTTTTAACTTTTACTATAAAGCCATGCAATGCAAAAACTATATTTACTTTCTCA	840							
Db	214266	ATGTTTTTTTAACTTTTACTATAAAGCCATGCAATGCAAAAACTATATTTACTTTCTCA	214207							
QY	841	AAATGGACCCCAAGAAACTTTGATTTACTGTTCTTAACTGCCATACTTAATTACAGGA	900							
Db	214206	AAATGGACCCCAAGAAACTTTGATTTACTGTTCTTAACTGCCATACTTAATTACAGGA	214147							
QY	901	ACTGTGCATCAGCTATTTATGATTTCTATAAGCTATTTTCCAGCAGAAATGAGATATTAACCC	960							
Db	214146	ACTGTGCATCAGCTATTTATGATTTCTATAAGCTATTTTCCAGCAGAAATGAGATATTAACCC	214087							
QY	961	AATGCTTTTGAATTTCTAGAAAAGTATAGTAATTTGTTTTCTAAGGTGGTTCAGCATCTA	1020							
Db	214086	AATGCTTTTGAATTTCTAGAAAAGTATAGTAATTTGTTTTCTAAGGTGGTTCAGCATCTA	214027							
QY	1021	CTCTTTTATCATTTACTTCAAAATGACATTTGCTAAAGACTGCATTTATTTTACTACTGTA	1080							
Db	214026	CTCTTTTATCATTTACTTCAAAATGACATTTGCTAAAGACTGCATTTATTTTACTACTGTA	213967							
QY	1081	ATTTCTCCACGACATAGCATTTATGTACATAGATGAGTGAACATTTATATCTCACATAGA	1140							
Db	213966	ATTTCTCCACGACATAGCATTTATGTACATAGATGAGTGAACATTTATATCTCACATAGA	213907							
QY	1141	GACATGCTTATATGGTTTTTATTTTAAATGAAATGCCAGTCCATTACACTGAATAAATAGA	1200							
Db	213906	GACATGCTTATATGGTTTTTATTTTAAATGAAATGCCAGTCCATTACACTGAATAAATAGA	213847							
QY	1201	ACTCAACTATTGCTTTTTCAGGGAATCATGGATAGGGTTGAAGAAGTTTACTATTATTG	1260							
Db	213846	ACTCAACTATTGCTTTTTCAGGGAATCATGGATAGGGTTGAAGAAGTTTACTATTATTG	213787							
QY	1261	TTTAAAAACAGCTTAGGGATTAATGTCCTCCATTTTATGAAGATTAAAAATGAAGGCTT	1320							
Db	213786	TTTAAAAACAGCTTAGGGATTAATGTCCTCCATTTTATGAAGATTAAAAATGAAGGCTT	213727							
QY	1321	TAATCAGCATTTGTAAGGAAATGAAATGGCTTTCTGATATGCTGTTTTTTTAGCCTAGGAG	1380							
Db	213726	TAATCAGCATTTGTAAGGAAATGAAATGGCTTTCTGATATGCTGTTTTTTTAGCCTAGGAG	213667							
QY	1381	TTAGAAATCCTAACTTCTTTATCCCTCTTCTCCAGAGGTTTTTTTTCTTGTGTATTAA	1440							
Db	213666	TTAGAAATCCTAACTTCTTTATCCCTCTTCTCCAGAGGTTTTTTTTCTTGTGTATTAA	213607							
QY	1441	ATTAACATTTTTTAAACGCAGATATTTTGTCAAGGGGCTTTGCATTCAAACTGCTTTTCC	1500							
Db	213606	ATTAACATTTTTTAAACGCAGATATTTTGTCAAGGGGCTTTGCATTCAAACTGCTTTTCC	213547							
QY	1501	AGGCTATACTCAGAAAGAAAGATAAAGTGTGATCTAAGAAAAAGTGAATGTTTTTAGGAA	1560							

QY 301 TCCCTGCTGGCTCTTTCTCCGGACCTACAGCAGCCAGAGGACTGATGTGTGCTTCC 360
DB 258563 TCCCTGCTGGCTCTTTCTCCGGACCTACAGCAGCCAGAGGACTGATGTGTGCTTCC 258504
QY 361 GTGATGTCTCTTCTGGCTTTTCAATGATGGCCATCCTTGGCATGAAATGCACCAGTGCACG 420
DB 258503 GTGATGTCTCTTCTGGCTTTTCAATGATGGCCATCCTTGGCATGAAATGCACCAGTGCACG 258444
QY 421 GGGGCAATGAGAAAGGTGAAGGCTCACATCTGCTGACGGCTGGAATCATCTTCATCATC 480
DB 258443 GGGGCAATGAGAAAGGTGAAGGCTCACATCTGCTGACGGCTGGAATCATCTTCATCATC 258384
QY 481 ACGGSCATGGTGGTCTCATCCCTGTGAGCTGGGTTGCCAATGCCATCATCAGAGATTTC 540
DB 258383 ACGGSCATGGTGGTCTCATCCCTGTGAGCTGGGTTGCCAATGCCATCATCAGAGATTTC 258324
QY 541 TATAACTCAATAGTGAATGTGCCCCAAAACCTGAGCTTGGAGAAGCTCTCTACTTAGGA 600
DB 258323 TATAACCCCAATAGTGAATGTGCCCCAAAACCTGAGCTTGGAGAAGCTCTCTACTTAGGA 258264
QY 601 TGGACCAAGGCTGGTGTGATTTGTTGGAGGAGCTCTGTTCTGCTGCGTTTTTGTGTC 660
DB 258263 TGGACCAAGGCTGGTGTGATTTGTTGGAGGAGCTCTGTTCTGCTGCGTTTTTGTGTC 258204
QY 661 AACGAAAAGAGCAGTAGCTACAGATACTCGATACCTTCCCATCGCACAAACCCAAAAAAGT 720
DB 258203 AACGAAAAGAGCAGTAGCTACAGATACTCGATACCTTCCCATCGCACAAACCCAAAAAAGT 258144
QY 721 TATCACACCGGAAAGAGTACCGAGCGTCTACTCCAGAAAGTCAGTATGTAGTGTGT 780
DB 258143 TATCACACCGGAAAGAGTACCGAGCGTCTACTCCAGAAAGTCAGTATGTAGTGTGT 258084
QY 781 ATGTTTTTTTAACTTTACTATAAAGCCATGCAAAATGACAAAAATCTATATTACTTTCTCA 840
DB 258083 ATGTTTTTTTAACTTTACTATAAAGCCATGCAAAATGACAAAAATCTATATTACTTTCTCA 258024
QY 841 AAATGACCCCAAGAAACTTTGATTTACTGTTCTTAACTGCTTAATCTTAAATACAGGA 900
DB 258023 AAATGACCCCAAGAAACTTTGATTTACTGTTCTTAACTGCTTAATCTTAAATACAGGA 257964
QY 901 ACTGTGCATCAGCTATTTATGATTTCTATAAGCTATTTTCCAGCAGAAATGAGATATTAAACCC 960
DB 257963 ACTGTGCATCAGCTATTTATGATTTCTATAAGCTATTTTCCAGCAGAAATGAGATATTAAACCC 257904
QY 961 AATGCTTTGATTTCTAGAAAGTATAGTAATTTGTTTTCTAAGTGTTCAAGCATCTA 1020
DB 257903 AATGCTTTGATTTCTAGAAAGTATAGTAATTTGTTTTCTAAGATGTTCAAGCATCTA 257844
QY 1021 CTCCTTTTATCATTTACTTCAAAATGACATTGTCTAAAGACTGCATTATTTTACTACTGTA 1080
DB 257843 CTCCTTTTATCATTTACTTCAAAATGACATTGTCTAAAGACTGCATTATTTTACCATTGTA 257784
QY 1081 ATTTCTCCACGACATAGCATATATGTACATAGATGAGTGTAACATTTATATCTCACATAGA 1140
DB 257783 ATTTCTCCACGACATAGCATATATGTACATAGATGAGTGTAACATTTATATCTCACATAGA 257724
QY 1141 GACATGCTTATATGGTTTTTATTTAAATGAAATGCCAGTCCATTACACTGAATAAATAGA 1200
DB 257723 GACATGCTTATATGGTTTTTATTTAAATGAAATGCCAATCCATTACACTGAATAAATAGA 257664
QY 1201 ACTCAACTATTTGCTTTTCCAGGAAATCATGGATAGGGTTGAAGAGGTTACTATTAAATTG 1260
DB 257663 ACTCAACTATTTGCTTTTCCAGGAAATCATGGATAGGGTTGAAGAGGTTACTATTAAATTG 257604
QY 1261 TTTAAAAACAGCTTAGGGATTAATGTCTCCATTTATAATGAAGATTAAAAATGAAGCTT 1320
DB 257603 TTTAAAAACAGCTTAGGGATTAATGTCTCCATTTATAATGAAGTTAAAAATGAAGCTT 257544
QY 1321 TAATCAGCATTTGTAAGGAAATGAATGGCTTTCTGATATGCTGTTTTTTAGCCCTAGGAG 1380
DB 257543 TAATCAGCATTTGTAAGGAAATGAATGGCTTTCTGATATGCTGTTTTTTAGCCCTAGGAG 257484

QY 1381 TTAGAAATCCTAACTTCTTTATCCTCTTCTCCAGAGGC-TTTTCTTTTCTTGTGTATTA 1439
DB 257483 TTAGAAATCCTAACTTCTTTTCTCTCTTCTCCAGAGGCTTTTCTTTTCTTGTGTATTA 257424
QY 1440 AATTAACATTTTAAACGCGAGATATTTTGTCAAGGGCTTTTGCAATCAAACCTGCTTTTC 1499
DB 257423 AATTAACATTTTAAACGCGAGAT-TTTTGTCAAGGGCTTTTGCAATCAAACCTGCTTTTC 257365
QY 1500 CAGGGCTATACTCAGAAAGAAAGATAAAAGTGTGATCTAAGAAAAAGTGTGTTTAGGA 1559
DB 257364 CAGGGCTATACTCAGAAAGAAAGATAAAAGTGTGATCTAAGAAAAAGTGTGTTTAGGA 257305
QY 1560 AAGTGAAAATATTTTGTGTTTGTATTTGAAGAAGAAATGATGCAATTTTGACAAGAAATCA 1619
DB 257304 AAGTGAAAATATTTTGTGTTTGTATTTGAAGAAGAAATGATGCAATTTTGACAAGAAATCA 257245
QY 1620 TATATGTATGGATATATTTTAAATAAGTATTTTGAGTACAGACTTTTGAGGTTTCATCAATAT 1679
DB 257244 TATATGTATGGATATATTTTAAATAAGGATTTTGAGTACAGACTTTTGAGGTTTCATCAATAT 257185
QY 1680 AAATAAAAGAGCAGAAAAATATGTCTTGTGTTTCAATTTGCTTACCAAAAAACAACAAC- 1738
DB 257184 AAATAAAAGAGCAGAAAAATATGTCTTGTGTTTCAATTTGCTTACCAAAAAACAACAACA 257125
QY 1739 AAAAAAAGTTCTCTTTGAGAACTTCACTGCTCTATGTGGTACCTGAGTCAAAATTG 1798
DB 257124 AAAAAAAGTTCTCTTTGAGAACTTCACTGCTCTATGTAGGTACCTGAGTCAAAATTG 257065
QY 1799 TCATTTTGTCTGTGAAAAATAAATTTCTTCTTGTGACCAATTTCTGTTAGTTTACTA 1858
DB 257064 TCATTTTGTCTGTGAAAAATAAATTTCTTCTTGTGACCAATTTCTGTTAGTTTACTA 257005
QY 1859 AAATCTGTAATACTGTATTTTCTGTTTATTTCCAAATTTGATGAAACTGACAATCCAAT 1918
DB 257004 AAATCTGTAATACTGTATTTTCTGTTTATTTCCAAATTTGATGAAACTGACAATCCAAT 256945
QY 1919 TTGAAAGTTTGTGTCGACGTCGTCTAGCTTAAATGAATGTGTTCTATTGCTTTATACA 1978
DB 256944 TTGAAAGTTTGTGTCGACATCTGTCTAGCTTAAATGAATTTGTTCTATTGCTTTATACA 256885
QY 1979 TTTATATTAATAAATTTGTACATTTTCTAATT 2010
DB 256884 TTTATATTAATAAATTTGTACATTTTCTAATT 256853

RESULT 8
BS000178/c

LOCUS Pan troglodytes chromosome 22 clone:RP43-042C06, map 22, complete
DEFINITION BS000178 176580 bp DNA linear PRI 07-OCT-2003
sequences.

ACCESSION BS000178
VERSION BS000178.1 GI:37537445

KEYWORDS

HTG.

SOURCE Pan troglodytes (chimpanzee)

ORGANISM Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE

1 The Chimpanzee Chromosome 22 Sequencing Consortium.

AUTHORS DNA sequence of chimpanzee chromosome 22 and its evolutionary
TITLE implications

JOURNAL

REFERENCE 2 (bases 1 to 176580)

AUTHORS Wang,S., Cai,Z., Wang,B., Zheng,H., Zhang,Y., Zhang,X., Zhu,G.,
Lu,G., Fu,G. and Chen,Z.

TITLE

Direct Submission

JOURNAL Submitted (26-MAY-2003) Shengyue Wang, Chinese National Human

Genome Center at Shanghai, Genomic Sequencing; No.250 BiBo Road,

Zhang Jiang HI-TECH Park, Shanghai 201203, CHINA

(E-mail:wangsy@chgc.sh.cn, URL:http://www.chgc.sh.cn,

Tel:86-21-50801919, Fax:86-21-50801922)

COMMENT The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
*Chinese National Human Genome Center at Shanghai, Shanghai, China;

Db 3004 ACTCAACTATTGCTTTTCAGGGAATCATGGATAGGATTGAAGAAGTTACTATTAATTG 29945
Qy 1261 TTTAAAAACAGCTTAGCGATTAAATGCTCTCCATTATATAATGAAGATTAAAAATGAAGGCTT 1320
Db 29944 TTTAAAAACAGCTTAGCGATTAAATGCTCTCCATTATATAATGAAGTTAAAAATGAAGGCTT 29885
Qy 1321 TAATCAGCATTTGTAAGGAAATGAATGCTTTCTGATATGCTGTTTCTTTAGCTAGGAG 1380
Db 29884 TAATCAGGATTGTAAGGAAATGAATGCTTTCTGATATGCTGTTTCTTTAGCTAGGAG 29825
Qy 1381 TTGAAATCCTAACTTTCTTTATCCTCTCTCCAGAGGC--TTTCTTTTCTTTGTTGTTATT 1438
Db 29824 TTGAAATCCTAACTT-TTTTCTCTCTCTCCAGAGGCTTTTCTTTTCTTTGTTGTTATT 29766
Qy 1439 AAATTAACATTTTAAACGCAGATATTTTGTCAAGGGCTTTTGCATTCAAACCTGCTTTT 1498
Db 29765 AAATTAACATTTTAAACGCAGATATTTTGTCAAGGGCTTTTGCATTCAAACCTGCTTTT 29706
Qy 1499 CCAGGGCTATACCTCAGAGAAAGATATAAAGTGTGATCTAAGAAAAAGTGATGCTTTTAGG 1558
Db 29705 CCAGGGCTATACCTCAGAGAAAGATATAAAGTGTGATCTAAGAAAAAGTGATGCTTTTAGG 29646
Qy 1559 AAAGTGAAAAATATTTTGTGTTTGTATTTGAAGAAGATGATGCAATTTTGACAGAAAAATC 1618
Db 29645 AAAGTGAAAAATATTTTGTGTTTGTATTTGAAGAAGATGATGCAATTTTGACAGAAAAATC 29586
Qy 1619 ATATATGATGATATATTTTAAATAAGTATTTGAGTACAGACTTTGAGGTTTTCATCAATA 1678
Db 29585 ATATATGATGATATATTTTAAATAAGTATTTGAGTACAGACTTTGAGGTTTTCATCAATA 29526
Qy 1679 TAAATAAAGAGCAGAAAAATATGCTTGGTTTTCATTTGCTTACCAAAAAACAACAAC 1738
Db 29525 TAAATAAAGAGCAGAAAAATATGCTTGGTTTTCATTTGCTTACCAAAAAACAACAAC 29466
Qy 1739 -AAAAAAGTTGCTCTTGAGAACTTCCCTGCTCCTATGCGGTACCTGAGTCAAAAATT 1797
Db 29465 AAAAAAAGTTGCTCTTGAGAACTTCCCTGCTCCTATGAGGTACCTGAGTCAAAAATT 29406
Qy 1798 GTCATTTTGTCTGTGAAAAATAAATTTCTTGTACCAATTTCTGTTTAGTTTACT 1857
Db 29405 GTCATTTTGTCTGTGAAAAATAAATTTCTTGTACCAATTTCTGTTTAGTTTACT 29346
Qy 1858 AAAATCTGTAATACTGTATTTTCTGTTTATTTCCAAATTTGATGAAACTGACAAATCCAA 1917
Db 29345 AAAATCTGTAATACTGTATTTTCTGTTTATTTCCAAATTTGATGAAACTGACAAATCCAA 29286
Qy 1918 TTTGAAAGTTTGTGTCGACGCTGTGCTAGCTTAAATGAATGTTCTATTGCTTTATAC 1977
Db 29285 TTTGAAAGTTTGTATAGACATCTGCTAGCTTAAATGAATTTTCTATTGCTTTATAC 29226
Qy 1978 ATTTATATTAATAAATTTGTACATTTTCTTAATT 2010
Db 29225 ATTTATATTAATAAATTTGTACATTTTCTTAATT 29193

RESULT 9
HSA250711 HSA250711 1931 bp DNA linear PRI 11-NOV-1999
LOCUS
DEFINITION Homo sapiens CLDN8 gene for claudin-8.
ACCESSION AJ250711
VERSION AJ250711.1 GI:6433859
KEYWORDS claudin-8; CLDN8 gene.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Keen,T.J. and Inglehearn,C.F.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1931)
AUTHORS Keen,T.J.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-1999) Keen T.J., Molecular Medicine Unit,

University of Leeds, Clinical Sciences Building, St James's
Hospital, Leeds, LS9 7TF, UNITED KINGDOM
FEATURES
source
Location/Qualifiers
1..1931
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
23..700
/gene="CLDN8"
23..700
/gene="CLDN8"
/codon_start=1
/product="claudin-8"
/protein_id="CAB60615.1"
/db_xref="GI:6433860"
/db_xref="GOA:P56748"
/db_xref="SWISS-PROT:P56748"
/translation="MATHALEIAGFLGGVMGTVAVTVPQWRVSAFIENNIIVVFE
NFEGLWMNCVRQANIRMQCKIYDSLALSPDLQAARGLMCAASVMSFLAFMMAILGM
KTRCTGDNEKVKAHILLTAGIIFIITGMVVLIPVSWANAIIRDYNSIVNVAKRE
LGEALYLGWTTALVLIIVGALFCCVFCNEKSSSRYISIPSHRTTQKSYHTGKSPSV
YRSQYV"
polyA_signal 1913..1918
ORIGIN

Query Match 96.0%; Score 1929.4; DB 9; Length 1931;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1930; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 75 CTAATACTCATCCAGAGGATAATGGCAACCCATGCTTAGAAAATCGCTGGGCTGTTTCT 134
Db 1 CTAATACTCATCCAGAGGATAATGGCAACCCATGCTTAGAAAATCGCTGGGCTGTTTCT 60
Qy 135 TGGTGGTGTGGAATGGTGGGACAGTGGTGTGCTGCTCATGCTCATGCTCAGTGGAGAGTGC 194
Db 61 TGGTGGTGTGGAATGGTGGGACAGTGGTGTGCTGCTCATGCTCAGTGGAGAGTGC 120
Qy 195 GGCCTTCATTGAAAAACAACATCTGTTTGTGAAACTTCTGGGAAGGACTGTGGATGAA 254
Db 121 GGCCTTCATTGAAAAACAACATCTGTTTGTGAAACTTCTGGGAAGGACTGTGGATGAA 180
Qy 255 TTGCGTGAGGAGGCTTAACATCAGGATGCAGTGCAAAATCTATGATTCCTGCTGGCTCT 314
Db 181 TTGCGTGAGGAGGCTTAACATCAGGATGCAGTGCAAAATCTATGATTCCTGCTGGCTCT 240
Qy 315 TTCTCCGACCTACAGGACGACAGGACTGATGTGCTGCTTCCGTGATGCTCTTCTT 374
Db 241 TTCTCCGACCTACAGGACGACAGGACTGATGTGCTGCTTCCGTGATGCTCTTCTT 300
Qy 375 GGCTTTCATGATGGCCATCCTTGGCATGAAATGCACAGGTGCACGGGGACAATGAGAA 434
Db 301 GGCTTTCATGATGGCCATCCTTGGCATGAAATGCACAGGTGCACGGGGACAATGAGAA 360
Qy 435 GGTGAAGGCTCACATTTCTGCTGACGGCTGGAATCATCTTTCATCATCAGGGCATGGTGT 494
Db 361 GGTGAAGGCTCACATTTCTGCTGACGGCTGGAATCATCTTTCATCATCAGGGCATGGTGT 420
Qy 495 GCTCATCCCTGTGAGCTGGTTGCCAATGCCATCATCAGAGATTTCTATAACTCAATAGT 554
Db 421 GCTCATCCCTGTGAGCTGGTTGCCAATGCCATCATCAGAGATTTCTATAACTCAATAGT 480
Qy 555 GAATGTTGCCAAAAACGTCGAGCTTGAGAGAGCTCTCTACTTAGGATGACCAACGCGACT 614
Db 481 GAATGTTGCCAAAAACGTCGAGCTTGAGAGAGCTCTCTACTTAGGATGACCAACGCGACT 540
Qy 615 GGTGCTGATTGTTGAGGAGCTCTGTTCTGCTGCGTTTTTGTGCAACGAAAAAGAGCAG 674
Db 541 GGTGCTGATTGTTGAGGAGCTCTGTTCTGCTGCGTTTTTGTGCAACGAAAAAGAGCAG 600
Qy 675 TAGCTACAGATACTCGATACCTTCCCATCGCACAAACCAAAAAAGTTATCACACCGGAAA 734

Db 601 TAGCTACAGATACTCGATACCTTCCCATCGCACACCCAAAAAGTTATCACACCGGAAA 660
Qy 735 GAAGTCACCGAGCGTCTACTCCAGAAGTCAGTATGTGTAGTGTGTATGTTTAACT 794
Db 661 GAAGTCACCGAGCGTCTACTCCAGAAGTCAGTATGTGTAGTGTGTATGTTTAACT 720
Qy 795 TTACTATAAAGCCATGCAAAATGACAAAAATCTATATACTTTCTCAAAATGGACCCCAA 854
Db 721 TTACTATAAAGCCATGCAAAATGACAAAAATCTATATACTTTCTCAAAATGGACCCCAA 780
Qy 855 GAACTTTGATTACTGTTCTTAACTGCCTAATCTTAATTACAGGAAGTGTGCATCAGCT 914
Db 781 GAACTTTGATTACTGTTCTTAACTGCCTAATCTTAATTACAGGAAGTGTGCATCAGCT 840
Qy 915 ATTTATGATTCTATAAGCTATTTTCAGCAGAATGAGATATTAACCCCAATGCTTTGATTGT 974
Db 841 ATTTATGATTCTATAAGCTATTTTCAGCAGAATGAGATATTAACCCCAATGCTTTGATTGT 900
Qy 975 TCTAGAAAGTATAGTAATTTGTTTCTAAGGTGGTTCAAGCATCTACTCTTTTATCAT 1034
Db 901 TCTAGAAAGTATAGTAATTTGTTTCTAAGGTGGTTCAAGCATCTACTCTTTTATCAT 960
Qy 1035 TACTTCAAAATGACATTGCTAAAGACTGCAATTATTTTACTACTGTAAATTTCTCCACGACA 1094
Db 961 TACTTCAAAATGACATTGCTAAAGACTGCAATTATTTTACTACTGTAAATTTCTCCACGACA 1020
Qy 1095 TAGCATTATGTACATAGATGAGTGTAAACATTTTATATCTCACATAGACATGCTTATATG 1154
Db 1021 TAGCATTATGTACATAGATGAGTGTAAACATTTTATATCTCACATAGACATGCTTATATG 1080
Qy 1155 GTTTTATTTAAATGAAATGCCAGTCCATTACACTGAATAAATAGAACTCAACTATTGCT 1214
Db 1081 GTTTTATTTAAATGAAATGCCAGTCCATTACACTGAATAAATAGAACTCAACTATTGCT 1140
Qy 1215 TTTCAGGGAAATCATGGATAGGGTTGAAGAAGTTACTATTAATTGTTTAAAAACAGCTT 1274
Db 1141 TTTCAGGGAAATCATGGATAGGGTTGAAGAAGTTACTATTAATTGTTTAAAAACAGCTT 1200
Qy 1275 AGGGATTAATGTCTCCATTTTATAATGAAGATTAAAAATGAAGCTTTAATCAGCATTGTA 1334
Db 1201 AGGGATTAATGTCTCCATTTTATAATGAAGATTAAAAATGAAGCTTTAATCAGCATTGTA 1260
Qy 1335 AAGGAAATGGAATGGCTTTCTGTATATGCTGTTTTTGTAGCCTAGGAGTAGAAATCCTAAC 1394
Db 1261 AAGGAAATGGAATGGCTTTCTGTATATGCTGTTTTTGTAGCCTAGGAGTAGAAATCCTAAC 1320
Qy 1395 TTCTTTTATCCTCTTCTCCAGAGGCTTTTTTTTTTTCTGTATTAAATTAACATTTTAA 1454
Db 1321 TTCTTTTATCCTCTTCTCCAGAGGCTTTTTTTTTTTCTGTATTAAATTAACATTTTAA 1380
Qy 1455 AACGCAGATATTTTGTCAAGGGCTTTTGCAATTCAACTGCTTTTCCAGGGCTATACCTAG 1514
Db 1381 AAGCAGATATTTTGTCAAGGGCTTTTGCAATTCAACTGCTTTTCCAGGGCTATACCTAG 1440
Qy 1515 AAGAAAGATAAAGTGTGATCTAAGAAAAAGTGTGTTTTTAGGAAAGTGAAAAATATTT 1574
Db 1441 AAGAAAGATAAAGTGTGATCTAAGAAAAAGTGTGTTTTTAGGAAAGTGAAAAATATTT 1500
Qy 1575 TGTTTTTGTATTTGAAGAAGATGATGCAATTTTGACAAGAAATCATATATGTATGGATAT 1634
Db 1501 TGTTTTTGTATTTGAAGAAGATGATGCAATTTTGACAAGAAATCATATATGTATGGATAT 1560
Qy 1635 ATTTTAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATATAAATAAAGAGCAGA 1694
Db 1561 ATTTTAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATATAAATAAAGAGCAGA 1620
Qy 1695 AAAATATGCTTGGTTTTTCAATTGCTTACCAAAAAACAACAATAAAGTTGTCTT 1754
Db 1621 AAAATATGCTTGGTTTTTCAATTGCTTACCAAAAAACAACAATAAAGTTGTCTT 1680
Qy 1755 TGAGAACTTCACCTGCTCCTATGTGGGTACCTGAGTCAAAATTTGTCATTTTGTCTGTG 1814
Db 1681 TGAGAACTTCACCTGCTCCTATGTGGGTACCTGAGTCAAAATTTGTCATTTTGTCTGTG 1740

Qy 1815 AAAAAATAAATTTCTTCTTGTACCAATTTCTGTTTAGTTTACTAAAAATCTGTAAATACTG 1874
Db 1741 AAAAAATAAATTTCTTCTTGTACCAATTTCTGTTTAGTTTACTAAAAATCTGTAAATACTG 1800
Qy 1875 TATTTTCTGTTTATTTCCAAATTTGATGAAACTGACAAATTTGAAAGTTTGTGTGCG 1934
Db 1801 TATTTTCTGTTTATTTCCAAATTTGATGAAACTGACAAATTTGAAAGTTTGTGTGCG 1860
Qy 1935 ACGTCTGTCTAGCTTAAATGAATGTGTTCTATTGTTTATACATTTTATAATAAATT 1994
Db 1861 ACGTCTGTCTAGCTTAAATGAATGTGTTCTATTGTTTATACATTTTATAATAAATT 1920
Qy 1995 GTACATTTTTC 2005
Db 1921 GTACATTTTTC 1931
RESULT 10
AX879004
LOCUS AX879004 1890 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 13909 from Patent EP1074617.
ACCESSION AX879004
VERSION AX879004.1 GI:40033740
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primers for synthesising full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 13909 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
source
1. 1890
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
150..851
/note="unnamed protein product"
/codon_start=1
/protein_id="CAE90333.1"
/db_xref="GI:40033741"
/translation="MATHALEIAGLFLLGVGMVGTAVTVPQWRVSFIENNVIFE
NFWGLMNCVROANIRMOCKIYDLSLLALSPDLQAARGMLCAASVMPFLAFMMAILGM
KCTRCTGDNEKVKAHILLTAGIIFIITGMVVLIPVSWANAIIRDYFNPIVNVAKRE
LGEALYLGTWTTALVLIVGALFCCVFCNEKSSSYRISIPSHRTTQKSYHTGKKSRL
QKSVGVVVYVFLTL"
CDS
ORIGIN
Query Match 89.7%; Score 1802; DB 6; Length 1890;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1831; Conservative 0; Mismatches 5; Indels 6; Gaps 2;
Qy 1 GGAAAAACTGTCTCTTCTGTGGCACAGAGAACCCCTGCTTCAAAGCAGAGTAGCAGTTC 60
Db 54 GGAAAAACTGTCTCTTCTGTGGCACAGAGAACCCCTGCTTCAAAGCAGAGTAGCAGTTC 113
Qy 61 CGGAGTCCAGCTGGCTAAACCTCATCCAGAGGATAATGGCAACCCATGCCCTTAGAAATC 120
Db 114 CGGAGTCCAGCTGGCTAAACCTCATCCAGAGGATAATGGCAACCCATGCCCTTAGAAATC 173
Qy 121 GCTGGCTGTCTTGTGGTGTGGAAATGGTGGGCACAGTGGTGTCACTGTGCATGCCT 180
Db 174 GCTGGCTGTCTTGTGGTGTGGAAATGGTGGGCACAGTGGTGTCACTGTGCATGCCT 233
Qy 181 CAGTGGAGAGTGTGGCCCTTCAATTGAAAAACAACATCGTGGTGTGTTTGAAGAACTTCTGGGAA 240
Db 234 CAGTGGAGAGTGTGGCCCTTCAATTGAAAAACAACATCGTGGTGTGTTTGAAGAACTTCTGGGAA 293
Qy 241 GGAAGTGTGATGAATTCGCTGAGGAGGCTAACATCAGGATGCAGTGCAGAAATCTATGAT 300

Query Match 89.7%; Score 1802; DB 6; Length 1890;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1831; Conservative 0; Mismatches 5; Indels 6; Gaps 2;

QY	1	GGAAAAA	CTGTTCTCTG	TGGCAGAGAA	CCCTGCTTCA	AAGCAGAG	AGTAGCAG	TTTC	60
DB	54	GGAAAAA	CTGTTCTCTG	TGGCAGAGAA	CCCTGCTTCA	AAGCAGAG	AGTAGCAG	TTTC	113
QY	61	CGGAGTCC	AGCTGGCTAAA	ACTCATCC	CAGAGGATA	TGGCAAC	CCCATGCTT	AGAAATC	120
DB	114	CGGAGTCC	AGCTGGCTAAA	ACTCATCC	CAGAGGATA	TGGCAAC	CCCATGCTT	AGAAATC	173
QY	121	GCTGGGCT	GTGTTCTTGG	TGGTGTGGA	ATGGTGGG	CACAGTGG	CTGTCACT	GTCAATC	180
DB	174	GCTGGGCT	GTGTTCTTGG	TGGTGTGGA	ATGGTGGG	CACAGTGG	CTGTCACT	GTCAATC	233
QY	181	CAGTGGAG	AGTGTGGGCT	TTCAATGAAA	CAACATCG	TGGTGTGGA	ATGGTGGG	CACTGTG	240
DB	234	CAGTGGAG	AGTGTGGGCT	TTCAATGAAA	CAACATCG	TGGTGTGGA	ATGGTGGG	CACTGTG	293
QY	241	GGACTGTG	ATGAATGCG	TGAGGCGCT	TAACATCAG	ATGCAAGT	GCAGTGC	AGTATGAT	300
DB	294	GGACTGTG	ATGAATGCG	TGAGGCGCT	TAACATCAG	ATGCAAGT	GCAGTGC	AGTATGAT	353
QY	301	TCCCTGCT	GGCTCTTCT	CCGGACCT	TACAGGC	CAGCCAGG	AGCTGATG	TGTGCTTCC	360
DB	354	TCCCTGCT	GGCTCTTCT	CCGGACCT	TACAGGC	CAGCCAGG	AGCTGATG	TGTGCTTCC	413
QY	361	GTGATGTC	CTTCTGCT	TTTCAATG	ATGGCCAT	CTTGGCAT	GAAATGC	ACAGGTGC	420
DB	414	GTGATGTC	CTTCTGCT	TTTCAATG	ATGGCCAT	CTTGGCAT	GAAATGC	ACAGGTGC	473
QY	421	GGGGACA	ATGAGAA	GGTGAAGG	CTCAGCT	CTGAGCGG	CTGGAAT	CTCATCATC	480
DB	474	GGGGACA	ATGAGAA	GGTGAAGG	CTCAGCT	CTGAGCGG	CTGGAAT	CTCATCATC	533
QY	481	ACGGGCAT	TGGTGTGCT	CATCCCTG	TGAGCTGG	GTGGCAAT	GCCATCAT	CAGATTTTC	540
DB	534	ACGGGCAT	TGGTGTGCT	CATCCCTG	TGAGCTGG	GTGGCAAT	GCCATCAT	CAGATTTTC	593
QY	541	TATAACT	CAATAGTGA	ATGTTGCC	CAAAACCT	GAGCTTGG	AGAGCTCT	CTACTTAGGA	600
DB	594	TATAACT	CAATAGTGA	ATGTTGCC	CAAAACCT	GAGCTTGG	AGAGCTCT	CTACTTAGGA	653
QY	601	TGGACCA	CGGCACTGG	TGTGCTGAT	TGTTGG	AGGAGCTCT	GTGCTGG	CTGCTTTC	660
DB	654	TGGACCA	CGGCACTGG	TGTGCTGAT	TGTTGG	AGGAGCTCT	GTGCTGG	CTGCTTTC	713
QY	661	AACGAAA	AGAGCAGT	AGCTACAG	ATCTCG	ATACCTTCC	CATCGCA	CAACCCAAA	720
DB	714	AACGAAA	AGAGCAGT	AGCTACAG	ATCTCG	ATACCTTCC	CATCGCA	CAACCCAAA	773
QY	721	TATCAC	ACCGGAA	AGAGTCA	CCGAGCGT	CTACTCC	AGAGTCA	GTGTGTGT	780
DB	774	TATCAC	ACCGGAA	AGAGTCA	CCGAGCGT	CTACTCC	AGAGTCA	GTGTGTGT	828
QY	781	ATGTTTT	TTTTTAA	CTTAA	AGCCATG	CAAAATG	CAAAATG	CTATATCT	840
DB	829	ATGTTTT	TTTTTAA	CTTAA	AGCCATG	CAAAATG	CAAAATG	CTATATCT	888
QY	841	AAATGG	ACCCCA	AGAACTT	GTATTTA	CTGTTCTT	AACTGCCT	TAATCTTA	900
DB	889	AAATGG	ACCCCA	AGAACTT	GTATTTA	CTGTTCTT	AACTGCCT	TAATCTTA	948
QY	901	ACTGTG	CATCAG	CTATTTA	TGATTTA	CTGTTCTT	AACTGCCT	TAATCTTA	960
DB	949	ACTGTG	CATCAG	CTATTTA	TGATTTA	CTGTTCTT	AACTGCCT	TAATCTTA	1008
QY	961	AATGCT	TTTGAT	TGTTCTA	GAAAGT	ATAGTA	ATTTGTTT	CTTAAG	1020
DB	1009	AATGCT	TTTGAT	TGTTCTA	GAAAGT	ATAGTA	ATTTGTTT	CTTAAG	1068

QY	1021	CTCTTTT	TATCA	TTTCA	AAATG	CATTTG	CTAAAG	ATGCTG	CAATAT	TTTTT	ACTACT	GTGA	1080
DB	1069	CTCTTTT	TATCA	TTTCA	AAATG	CATTTG	CTAAAG	ATGCTG	CAATAT	TTTTT	ACTACT	GTGA	1128
QY	1081	ATTCTCC	ACGAC	ATAG	CATTAT	GTAC	ATAG	ATGAGT	GTAAAC	ATTATAT	CTCAC	ATAGA	1140
DB	1129	ATTCTCC	ACGAC	ATAG	CATTAT	GTAC	ATAG	ATGAGT	GTAAAC	ATTATAT	CTCAC	ATAGA	1188
QY	1141	GACATG	CTTAT	ATGTTT	TATTTA	AAATG	AAATGCC	AGTCC	ATTAC	ACTG	AAATAA	ATAGA	1200
DB	1189	GACATG	CTTAT	ATGTTT	TATTTA	AAATG	AAATGCC	AGTCC	ATTAC	ACTG	AAATAA	ATAGA	1248
QY	1201	ACTCAAC	TATG	CTTTT	CAGG	AAATC	ATGAT	AGGTTG	GAAG	AGGTT	TACTAT	TAATTG	1260
DB	1249	ACTCAAC	TATG	CTTTT	CAGG	AAATC	ATGAT	AGGTTG	GAAG	AGGTT	TACTAT	TAATTG	1308
QY	1261	TTTAAAA	CAGCTT	AGG	ATTAA	TGCTC	CCATT	TATA	TGAAG	ATTAA	ATGAAG	GCTT	1320
DB	1309	TTTAAAA	CAGCTT	AGG	ATTAA	TGCTC	CCATT	TATA	TGAAG	ATTAA	ATGAAG	GCTT	1368
QY	1321	TAATCAG	CATTG	TAAAG	AAAT	TGAAT	GGCTT	CTG	ATATG	CTGTTTT	TAGCCT	AGGAG	1380
DB	1369	TAATCAG	CATTG	TAAAG	AAAT	TGAAT	GGCTT	CTG	ATATG	CTGTTTT	TAGCCT	AGGAG	1428
QY	1381	TTAGAA	ATCCTA	CTTCT	TATCCT	CTTCTC	CCAGAG	GC -	TTTTT	TTTTT	CTTGTG	TATTA	1439
DB	1429	TTAGAA	ATCCTA	CTTCT	TATCCT	CTTCTC	CCAGAG	GC	TTTTT	TTTTT	CTTGTG	TATTA	1488
QY	1440	AATTAAC	ATTTT	TAAAA	CGCAG	ATATTT	TGTCA	AGGG	CTTTG	CATTC	AAACTG	CTTTTC	1499
DB	1489	AATTAAC	ATTTT	TAAAA	CGCAG	ATATTT	TGTCA	AGGG	CTTTG	CATTC	AAACTG	CTTTTC	1548
QY	1500	CAGGGCT	ATACT	CAAG	AAAG	ATAAA	AGTGT	GATCTA	AGAA	AAAGT	GATGTTT	TAGGA	1559
DB	1549	CAGGGCT	ATACT	CAAG	AAAG	ATAAA	AGTGT	GATCTA	AGAA	AAAGT	GATGTTT	TAGGA	1608
QY	1560	AAGTGAA	ATATTT	TGTTTT	TGTTT	GTTG	TAATG	AAAG	AAATG	ATGATG	CAAG	AAATCA	1619
DB	1609	AAGTGAA	ATATTT	TGTTTT	TGTTT	GTTG	TAATG	AAAG	AAATG	ATGATG	CAAG	AAATCA	1668
QY	1620	TATATG	TATGG	ATATAT	TTTAA	TAAAT	TTGAG	TATTT	GAGT	ACAG	CTTTG	AGGTTT	1679
DB	1669	TATATG	TATGG	ATATAT	TTTAA	TAAAT	TTGAG	TATTT	GAGT	ACAG	CTTTG	AGGTTT	1728
QY	1680	AAATAAA	AGAGC	AGAAAA	ATATG	CTTGG	TTTTT	CATTTG	CTTAC	CAAAAA	CAACAA	CA	1739
DB	1729	AAATAAA	AGAGC	AGAAAA	ATATG	CTTGG	TTTTT	CATTTG	CTTAC	CAAAAA	CAACAA	CA	1788
QY	1740	AAAAA	AGTTG	TCCTT	TGAG	AACTT	CACTG	CTCCT	ATG	TGGG	TACCTG	AGTCAA	1799
DB	1789	AAAAA	AGTTG	TCCTT	TGAG	AACTT	CACTG	CTCCT	ATG	TGGG	TACCTG	AGTCAA	1848
QY	1800	CATTTT	TGTTCT	GTGAAA	AAATA	TTTCT	CTTCT	TGTTAC	CAAT	1841			
DB	1849	CATTTT	TGTTCT	GTGAAA	AAATA	TTTCT	CTTCT	TGTTAC	CAAT	1890			

RESULT 12
AK022269
LOCUS
DEFINITION
Homoc sapiens cDNA FLJ12207 fis, clone MAMMA100956, highly similar to Homo sapiens CLDN8 gene for claudin-8.
ACCESSION
AK022269
VERSION
AK022269.1 GI:10433628
KEYWORDS
oligo capping; fis (full insert sequence).
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
1
AUTHORS
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sato, H., Sugano, S., Shiratori, A., Sudo, H., Wagatsuma, M., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,

Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
NEDO human cDNA sequencing project

Unpublished
2 (bases 1 to 1890)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.

FEATURES
source
1..1890
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MAMMA1000956"
/tissue type="Mammary gland"
/clone_lib="MAMMA1"
/note="cloning vector: pME18SFL3"
795^796
/note="5 bases segment is present in AJ250711 , AP001707.1
and AP000884.1."

ORIGIN
Query Match 89.7%; Score 1802; DB 9; Length 1890;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1831; Conservative 0; Mismatches 5; Indels 6; Gaps 2;

QY 1 GGAAAAAAGTCTCTCTCTGTGGCAGAGAAACCCCTGCTTCAAAGCAGAAGTAGCAGTTC 60
Db |||||
54 GGAAAAAAGTCTCTCTCTGTGGCAGAGAAACCCCTGCTTCAAAGCAGAAGTAGCAGTTC 113
QY 61 CGAGTCCAGCTGGCTTAAACTCATCCAGAGGATAATGGCAACCCATGCCCTTAGAAATC 120
Db |||||
114 CGAGTCCAGCTGGCTTAAACTCATCCAGAGGATAATGGCAACCCATGCCCTTAGAAATC 173
QY 121 GCTGGCTGTTCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 180
Db |||||
174 GCTGGCTGTTCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 233
QY 181 CAGTGGAGAGTGTGGCTTCAATTGAAACAAACATCGTGGTGGTGGTGGTGGTGGTGGTGGT 240
Db |||||
234 CAGTGGAGAGTGTGGCTTCAATTGAAACAAACATCGTGGTGGTGGTGGTGGTGGTGGTGGT 293
QY 241 GGACTGTGGATGAATTGGCTGAGGCAGGCTAAACATCAGGATGCGAGTGCAGTGCAGTGCAT 300
Db |||||
294 GGACTGTGGATGAATTGGCTGAGGCAGGCTAAACATCAGGATGCGAGTGCAGTGCAGTGCAT 353
QY 301 TCCCTGCTGGCTCTTCTCCGGACCTACAGGAGCCAGAGGACTGATGTGTGCTGCTTCC 360
Db |||||
354 TCCCTGCTGGCTCTTCTCCGGACCTACAGGAGCCAGAGGACTGATGTGTGCTGCTTCC 413
QY 361 GTGATGTGCTTCTTGGCTTTCATGATGGCCATCCTTGGCATGAAATGCACAGGTGCACG 420
Db |||||
414 GTGATGCTTCTTGGCTTTCATGATGGCCATCCTTGGCATGAAATGCACAGGTGCACG 473
QY 421 GGGGACAATGAGAAGGTGAAGGCTCACATTCGCTGACGGCTGGAATCATCTTCATCATC 480
Db |||||
474 GGGGACAATGAGAAGGTGAAGGCTCACATTCGCTGACGGCTGGAATCATCTTCATCATC 533
QY 481 ACGGGCATGGTGGTGTCTATCCCTGTGAGCTGGGTGCCAATGCCATCATCAGAGATTC 540
Db |||||
534 ACGGGCATGGTGGTGTCTATCCCTGTGAGCTGGGTGCCAATGCCATCATCAGAGATTC 593

QY 541 TATRAACTCAATAGTGAATGTTGCCCAAAAAACGCTGAGCTTGGAGAAGCTCTCTACTTAGGA 600
Db |||||
594 TATAACCCCAATAGTGAATGTTGCCCAAAAAACGCTGAGCTTGGAGAAGCTCTCTACTTAGGA 653
QY 601 TGGACCACGGCACTGGTGTCTGATTGTTGGAGGAGCTCTGTCTCTGCTGCGTCTTTTGTGTC 660
Db |||||
654 TGGACCACGGCACTGGTGTCTGATTGTTGGAGGAGCTCTGTCTCTGCTGCGTCTTTTGTGTC 713
QY 661 AACGAAAAGAGCAGTAGCTACAGATACTCGATACCTTCCCATCGCACAAACCCAAAAAAGT 720
Db |||||
714 AACGAAAAGAGCAGTAGCTACAGATACTCGATACCTTCCCATCGCACAAACCCAAAAAAGT 773
QY 721 TATCACACGGGAAAGAGTCACCGAGCGTCTACTCCAGAAAGTCAGTATGTGTAGTTGTGT 780
Db |||||
774 TATCACACGGGAAAGAGTCA-----CGTCTACTCCAGAAAGTCAGTATGTGTAGTTGTGT 828
QY 781 ATGTTTTTTTAACTTTTACTATAAAGCCATGCAAAATGACAAAAATCTATATTACTTTCTCA 840
Db |||||
829 ATGTCTTTTAACTTTTACTATAAAGCCATGCAAAATGACAAAAATCTATATTACTTTCTCA 888
QY 841 AAATGGACCCCAAGAAACTTTGATTTACTGTCTCTTAACTGCCTTAATCTTAATTACAGGA 900
Db |||||
889 AAATGGACCCCAAGAAACTTTGATTTACTGTCTTAACTGCCTTAATCTTAATTACAGGA 948
QY 901 ACTGTGCATCAGCTATTTATGATTTCTATAAGCTATTTTTCAGCAGAAATGAGATATTTAAACCC 960
Db |||||
949 ACTGTGCATCAGCTATTTATGATTTCTATAAGCTATTTTTCAGCAGAAATGAGATATTTAAATCC 1008
QY 961 AATGCTTTTGATTTCTAGAAAGTATAGTAAATTTTCTAAGGTGGTTCAGAGCATCTA 1020
Db |||||
1009 AATGCTTTTGATTTCTAGAAAGTATAGTAAATTTTCTAAGGTGGTTCAGAGCATCTA 1068
QY 1021 CTCCTTTTATCATTTACTTCAAAATGACATTTGCTAAAGACTGCATTAATTTTACTACTGTA 1080
Db |||||
1069 CTCCTTTTATCATTTACTTCAAAATGACATTTGCTAAAGACTGCATTAATTTTACTACTGTA 1128
QY 1081 ATTTCTCCAGACATAGCATTTATGTACATAGATGAGTGTAAACATTTATPATCTCACATAGA 1140
Db |||||
1129 ATTTCTCCAGACATAGCATTTATGTACATAGATGAGTGTAAACATTTATPATCTCACATAGA 1188
QY 1141 GACATGCTTATATGGTTTTTAAATGAAATGCCAGTCCATTACACTGAATAAATAGA 1200
Db |||||
1189 GACATGCTTATATGGTTTTTAAATGAAATGCCAGTCCATTACACTGAATAAATAGA 1248
QY 1201 ACTCAACTATTGCTTTTCAGGGAATCATGGATAGGGTTGAAGAAGGTACTATTAAATTG 1260
Db |||||
1249 ACTCAACTATTGCTTTTCAGGGAATCATGGATAGGGTTGAAGAAGGTACTATTAAATTG 1308
QY 1261 TTTAAAAACAGCTTAGGGATTAATGTCTCCATTTATAATGAAGATTAAAAATGAAGGCTT 1320
Db |||||
1309 TTTAAAAACAGCTTAGGGATTAATGTCTCCATTTATAATGAAGATTAAAAATGAAGGCTT 1368
QY 1321 TAATCAGCATTTGTAAAGGAAATTGAATGGCTTTCTGATATGCTGTTTTTTAGCCTAGGAG 1380
Db |||||
1369 TAATCAGCATTTGTAAAGGAAATTGAATGGCTTTCTGATATGCTGTTTTTTAGCCTAGGAG 1428
QY 1381 TTAGAAATCCTAACTTCTTTATCCTCTTCTCCAGAGGC- TTTTTTTTCTTGTGTATTA 1439
Db |||||
1429 TTAGAAATCCTAACTTCTTTATCCTCTTCTCCAGAGGCCTTTTTTTTCTTGTGTATTA 1488
QY 1440 AATTAACATTTTAAAAACGCAGATAATTTGTCAAGGGCTTTTGCATTCAAACCTGCTTTTC 1499
Db |||||
1489 AATTAACATTTTAAAAAGCAGATAATTTGTCAAGGGCTTTTGCATTCAAACCTGCTTTTC 1548
QY 1500 CAGGGCTATCTCAGAAAGAAAGATAAAAGTGTGATCTAAGAAAAAGTATGCGTTTATAGGA 1559
Db |||||
1549 CAGGGCTATCTCAGAAAGAAAGATAAAAGTGTGATCTAAGAAAAAGTATGCGTTTATAGGA 1608
QY 1560 AAGTGAAAAATATTTTGTGTTTGTATTTGAAGAAGATGATGCATTTTGAAGAATAATCA 1619
Db |||||
1609 AAGTGAAAAATATTTTGTGTTTGTATTTGAAGAAGATGATGCATTTTGAAGAATAATCA 1668
QY 1620 TATATGTATGGATATATTTTAAATAGTATTTGAGTACAGACTTTTGAGGTTTTCATCAATAT 1679

Db 1669 TATATGATGATATATTTTAAATAAGTATTGTGATACAGACTTTGAGGTTTCATCAATAT 1728
QY 1680 AAATAAAAGAGCAGAAAAATATGCTTGGTTTTCATTGCTTACCRAAAAAACAACA 1739
Db 1729 AAATAAAAGAGCAGAAAAATATGCTTGGTTTTCATTGCTTACCRAAAAAACAACA 1788
QY 1740 AAAAAAGTTGCTCTTGAGAACTTCACCTGCTCCTATGTGGGTACCTGAGTCAAAATTGT 1799
Db 1789 AAAAAAGTTGCTCTTGAGAACTTCACCTGCTCCTATGTGGGTACCTGAGTCAAAATTGT 1848
QY 1800 CATTTTGTCTGTGAAAAATAAATTTCTCTTGTGACCATT 1841
Db 1849 CATTTTGTCTGTGAAAAATAAATTTCTCTTGTGACCATT 1890

RESULT 13
BC058004
LOCUS
DEFINITION Homo sapiens claudin 8, mRNA (cDNA clone MGC:61830 IMAGE:4606880), complete cds.
ACCESSION BC058004
VERSION BC058004.1 GI:34783654
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1700)

REFERENCE
AUTHORS Strausberg,R.L., Feingold,B.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.P., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Young,A.C., Shevchenko,Y., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalish,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 1700)

REFERENCE
JOURNAL Strausberg,R.
MEDLINE Direct Submission
PUBMED Submitted (08-SEP-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
TITLE DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;
JOURNAL Web site: <http://www.nisc.nih.gov/>
AUTHORS Contact: nisc_mgc@hgri.nih.gov
TITLE Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Green,K., Brinkley,C., Brooks,S.,

Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 48 Row: d Column: 13
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21314655.

FEATURES
Location/Qualifiers
1..1700
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:61830 IMAGE:4606880"
/tissue_type="Kidney"
/clone_lib="NIH_MGC_75"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
1..1700
/gene="CLDN8"
/db_xref="LocusID:9073"
76..753
/codon_start=1
/product="CLDN8 protein"
/protein_id="AAH58004.1"
/db_xref="GI:34783655"
/db_xref="LocusID:9073"
/translation="NATHALEIAGFLGGVMGTAVTVMQWRVSFAFIENNIIVFE
NFEGLMNCVQRQANRMQCKIYDSLALSPDLQAAAGLMLCAASVMSFLAFMAILGM
KCTRTGDNEKVKRAHLLTAGIIFIITGMVLLIPVSWANAIIRDYNSIVNVAQKRE
LGEALYLGWTTTALVLIVGALFCCVFCNEKSSSYRYSIPSHRTTQKSYHTGKKSPV
YSRSQYV"
misc_feature
160..621
/note="PMP22_Claudin; Region: PMP-22/EMP/MP20/Claudin family"
/db_xref="CDD:pfam00822"

ORIGIN
Query Match 83.4%; Score 1675.8; DB 9; Length 1700;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1677; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 22 GGCACAGAGAACCTGCTTCAAGACAGAGTAGCAGTTCGGAGTCCAGCTGGCTAAAC 81
Db 1 GGGACAGAGAACCTGCTTCAAGACAGAGTAGCAGTTCGGAGTCCAGCTGGCTAAAC 60
QY 82 TCATCCCAGAGGATAATGGCAACCCATGCCCTAGAAATCGCTGGGCTGTTTCTTGGTGT 141
Db 61 TCATCCCAGAGGATAATGGCAACCCATGCCCTAGAAATCGCTGGGCTGTTTCTTGGTGT 120
QY 142 GTTGAATGTTGGGCACAGTGGCTGTCACTGTCACTGCCTCAGTGGAGAGTGTGCGGCTTC 201
Db 121 GTTGAATGTTGGGCACAGTGGCTGTCACTGTCACTGCCTCAGTGGAGAGTGTGCGGCTTC 180
QY 202 ATTGAAAACAACATCGTGGTTTTTTGAAAACCTTCTGGGAAGGACTGTGGATGAATTGGGTG 261
Db 181 ATTGAAAACAACATCGTGGTTTTTTGAAAACCTTCTGGGAAGGACTGTGGATGAATTGGGTG 240
QY 262 AGGCAGGCTAACATCAGGATGCAGTGCAAAATCTATGATTCCTGCTGGCTCTTTCTCCG 321
Db 241 AGGCAGGCTAACATCAGGATGCAGTGCAAAATCTATGATTCCTGCTGGCTCTTTCTCCG 300
QY 322 GACCTACAGGCAGCCAGAGACTGATGTGTGCTTCCGTGATGTCTTCTTGGCTTTC 381
Db 301 GACCTACAGGCAGCCAGAGACTGATGTGTGCTTCCGTGATGTCTTCTTGGCTTTC 360
QY 382 ATGATGGCCATCCTTGGCATGAAATGCACAGGTGCACGGGGACAATGAGAAGTGAAG 441
Db 361 ATGATGGCCATCCTTGGCATGAAATGCACAGGTGCACGGGGACAATGAGAAGTGAAG 420

QY 442 GCTCACATTCTGCTGACGGCTGGAATCATCTTTCATCATCAGGGGCATGGTGGTCTCATC 501
Db 421 GCTCACATTCTGCTGACGGCTGGAATCATCTTTCATCATCAGGGGCATGGTGGTCTCATC 480
QY 502 CCTGTGAGCTGGGTGCGCAATGCCATCATCAGAGATTTCATATAACTCAATAGTGAATGTT 561
Db 481 CCTGTGAGCTGGGTGCGCAATGCCATCATCAGAGATTTCATATAACTCAATAGTGAATGTT 540
QY 562 GCCCAAAAACGTGAGCTTGGAGAAGCTCTCTACTTAGGATGGACCA CGGCACTGGTGTG 621
Db 541 GCCCAAAAACGTGAGCTTGGAGAAGCTCTCTACTTAGGATGGACCA CGGCACTGGTGTG 600
QY 622 ATTGTTGGAGGAGCTCTGTTCTGCTGCGTTTTTTTGTGTCACGAAAAGAGAGTAGCTAC 681
Db 601 ANTGTTGGAGGAGCTCTGTTCTGCTGCGTTTTTTTGTGTCACGAAAAGAGAGTAGCTAC 660
QY 682 AGATACTCGATACCTTCCCATCGCACAAACCCAAAAAGTTATCACACCGGAAGAAGTCA 741
Db 661 AGATACTCGATACCTTCCCATCGCACAAACCCAAAAAGTTATCACACCGGAAGAAGTCA 720
QY 742 CCGAGCGTCTACTCCAGAAAGTCAGTATGTGTAGTTGTGTATGTTTTTTAACTTTACTAT 801
Db 721 CCGAGCGTCTACTCCAGAAAGTCAGTATGTGTAGTTGTGTATGTTTTTTAACTTTACTAT 780
QY 802 AAAGCCATGCAAAATGACAAAAATCTATATTACTTTCTCAAAATGGACCCCAAGAAACTT 861
Db 781 AAAGCCATGCAAAATGACAAAAATCTATATTACTTTCTCAAAATGGACCCCAAGAAACTT 840
QY 862 TGATTTACTGTTCTTAACTGCCTAATCTTAATTAACAGGAAGTGTGCATCAGCTATTATTG 921
Db 841 TGATTTACTGTTCTTAACTGCCTAATCTTAATTAACAGGAAGTGTGCATCAGCTATTATTG 900
QY 922 ATTCATAGCTATTTCTAGCAGAGAATGAGATATTAAACCCCAATGCTTTTGATTGTTCTAGAA 981
Db 901 ATTCATAGCTATTTCTAGCAGAGAATGAGATATTAAACCCCAATGCTTTTGATTGTTCTAGAA 960
QY 982 AGTATAGTAATTTGTTTCTAAGGTGGTTCAGCATCTACTCTTTTTTATCATTTACTTCA 1041
Db 961 AGTATAGTAATTTGTTTCTAAGGTGGTTCAGCATCTACTCTTTTTTATCATTTACTTCA 1020
QY 1042 AAATGACATTGCTAAGACTGCATTATTTTACTACTGTAAATTTCTCCACGACATAGCAAT 1101
Db 1021 AAATGACATTGCTAAGACTGCATTATTTTACTACTGTAAATTTCTCCACGACATAGCAAT 1080
QY 1102 ATGTACATAGATGAGTGAACATTTTATATCTCACATAGACATGCTTATATGGTTTTAT 1161
Db 1081 ATGTACATAGATGAGTGAACATTTTATATCTCACATAGACATGCTTATATGGTTTTAT 1140
QY 1162 TTAAAAATGAAATGCCAGTCCATTACACTGAATAAATAGAACTCAACTATTGCTTTTCAGG 1221
Db 1141 TTAAAAATGAAATGCCAGTCCATTACACTGAATAAATAGAACTCAACTATTGCTTTTCAGG 1200
QY 1222 GAAATCATGGATAGGTTGAAGAAGTTACTATTATTAATGTTTAAAAACAGCTTAGGGATT 1281
Db 1201 GAAATCATGGATAGGTTGAAGAAGTTACTATTATTAATGTTTAAAAACAGCTTAGGGATT 1260
QY 1282 AATGTCCTCCATTATATAATGAAGATTAAATGAAGGCTTTAATCAGCATTTGTAAGGAAA 1341
Db 1261 AATGTCCTCCATTATATAATGAAGATTAAATGAAGGCTTTAATCAGCATTTGTAAGGAAA 1320
QY 1342 TTGAATGGCTTTCTGATATGCTGTTTTTTAGCCTAGGAGTTAGAAATCCTAACTTCTTTA 1401
Db 1321 TTGAATGGCTTTCTGATATGCTGTTTTTTAGCCTAGGAGTTAGAAATCCTAACTTCTTTA 1380
QY 1402 TCCTCTTCTCCAGAGGCTTTTTTTTTCTTGTGTATTAAATTAACATTTTTTAAACGCAG 1461
Db 1381 TCCTCTTCTCCAGAGGCTTTTTTTTTCTTGTGTATTAAATTAACATTTTTTAAACGCAG 1440
QY 1462 ATATTTTGTCAAGGGGCTTTGCATTCAAACTGCTTTTCCAGGGCTATACTCAGAAGAAAG 1521
Db 1441 ATATTTTGTCAAGGGGCTTTGCATTCAAACTGCTTTTCCAGGGCTATACTCAGAAGAAAG 1500
QY 1522 ATAAAAAGTGTGATCTAAGAAAAAAGTGATGCTGTTTTTAGGAAAAGTGAAAAATATTTTTGTTTTT 1581

Db 1501 ATAAAAAGTGTGATCTAGAAAAAGTGATGGTTTTAGGAAAAGTGAATAATATTTTTGTTTTT 1560
QY 1582 GTATTGAAAGAAATGATGCATTTTGACAGAAATCATATATGATGGATATATTTTAA 1641
Db 1561 GTATTTGAAAGAAATGATGCATTTTGACAGAAATCATATATGATGGATATATTTTAA 1620
QY 1642 TAAGTATTGAGTACAGACTTTGAGGTTTCATCAATATAATAAAGAGCAGAAAAATA 1700
Db 1621 TAAGTATTGAGTACAGACTTTGAGGTTTCATCAATATAATAAAGAGCAGAAAAATA 1679

RESULT 14
BC020866
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT

BC020866 1835 bp mRNA linear PRI 06-OCT-2003
Homo sapiens claudin 8, mRNA (cDNA clone MGC:24067 IMAGE:4594155), complete cds.
BC020866 GI:18089189
MGC.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 1835)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, A.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 1835)
Strausberg, R.
Direct Submission
Submitted (03-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 36 Row: e Column: 23
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21314655.

FEATURES	Location/Qualifiers	
	source	1..1835
	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/db_xref="taxon:9606"	
	/clone="MGC:24067 IMAGE:4594155"	
	/tissue_type="kidney"	
	/clone_lib="NIH MGC_75"	
	/lab_host="DH10B"	
	/note="Vector: pDNR-LIB"	
	1..1835	
	/gene="CLDN8"	
gene	/db_xref="LocusID:9073"	
CDS	110..787	
	/codon_start=1	
	/product="claudin 8"	
	/protein_id="AAH20866.1"	
	/db_xref="GI:18089190"	
	/db_xref="LocusID:9073"	
	/translation="MATHALEIAGLFLGGVGMVGTVAVTVMPOWRVSAPIENNIVFE	
	NFWEGLMNCVRQANIRMQCKIYDSLALS PDLQARGLMCAASVMSFLAFMMAILGM	
	KCTRCTGDNEKVKAHILLTAGIIFIITGMVLI PVSWANAI IIRDFNSIVNVAQKRE	
	LGEALYLGWTTALVILVGGALFCCVPCNEKSSYRISIPSHRITTKYSYHTGKKSPSV	
misc_feature	YRSQYV"	
	194..655	
	/note="PMP22_Claudin; Region: PMP-22/EMP/MP20/Claudin	
	family"	
ORIGIN	/db_xref="CDD:pfam00822"	
	Query Match 77.1%; Score 1550.4; DB 9; Length 1835;	
	Best Local Similarity 88.8%; Pred. No. 9.5e-301;	
	Matches 1785; Conservative 0; Mismatches 1; Indels 224; Gaps 1;	
QY	1	GGAAAACTGTTCTCTTCTGTGGCACAGAGAACCCCTGCTTCAAAGCAGAGTAGCAGTTC 60
Db	14	GGAAAACTGTTCTCTTCTGTGGCACAGAGAACCCCTGCTTCAAAGCAGAGTAGCAGTTC 73
QY	61	CGGAGTCCAGCTGGCTAAAACTCATCCCAGAGGATAATGGCAACCCCATGCCCTTAGAAATC 120
Db	74	CGGAGTCCAGCTGGCTAAAACTCATCCCAGAGGATAATGGCAACCCCATGCCCTTAGAAATC 133
QY	121	GCTGGGTGTTTCTTGGTGGTGTGGAAATGGTGGGCACAGTGGCTGTCACTGTCTATGCCT 180
Db	134	GCTGGGTGTTTCTTGGTGGTGTGGAAATGGTGGGCACAGTGGCTGTCACTGTCTATGCCT 193
QY	181	CAGTGGAGAGTGTGGGCCCTTCAATTGAAAAACAACATCGTGGTGTGTTTGAAAACTTCGGGAA 240
Db	194	CAGTGGAGAGTGTGGGCCCTTCAATTGAAAAACAACATCGTGGTGTGTTTGAAAACTTCGGGAA 253
QY	241	GGACTGTGGATGAATTGCGTGAGGCAGGCTAAACATCAGGATGCGAGTGCAGGATCTATGAT 300
Db	254	GGACTGTGGATGAATTGCGTGAGGCAGGCTAAACATCAGGATGCGAGTGCAGGATCTATGAT 313
QY	301	TCCCTGCTGGCTCTTTCTCCGACCTACAGCAGCCAGAGGACTGATGTGTGCTGCTTCC 360
Db	314	TCCCTGCTGGCTCTTTCTCCGACCTACAGCAGCCAGAGGACTGATGTGTGCTGCTTCC 373
QY	361	GTGATGCTCTTCTTGGCTTTTCATGATGGCCATCCTTGGCATGAAATGCACCAGGTGCACG 420
Db	374	GTGATGCTCTTCTTGGCTTTTCATGATGGCCATCCTTGGCATGAAATGCACCAGGTGCACG 433
QY	421	GGGGACAATGAGAAGGTGAAGGCTCACATTTCTGCTGACGGCTGGAATCATCTTCATCATC 480
Db	434	GGGGACAATGAGAAGGTGAAGGCTCACATTTCTGCTGACGGCTGGAATCATCTTCATCATC 493
QY	481	ACGGGATGGTGGTGTCTATCCCTGTGAGCTGGGTGGCAATGCCATCATCAGAGATTTC 540
Db	494	ACGGGATGGTGGTGTCTATCCCTGTGAGCTGGGTGGCAATGCCATCATCAGAGATTTC 553
QY	541	TATAACTCAATAGTGAATGTTGCCCAAAAACGCTGAGCTTGGAGAAGCTCTCTACTTAGGA 600
Db	554	TATAACTCAATAGTGAATGTTGCCCAAAAACGCTGAGCTTGGAGAAGCTCTCTACTTAGGA 613

QY	601	TGGACCACGGCACTGGTGTCTGATGTTGGAGGAGCTCTGTTCTGCTGCGTTTTTTTGTTC 660
Db	614	TGGACCACGGCACTGGTGTCTGATGTTGGAGGAGCTCTGTTCTGCTGCGTTTTTTTGTTC 673
QY	661	AACGAAAAAGAGCAGTAGCTACAGATACTCGATACCTTCCCATCGCAACCCCAAAAAAGT 720
Db	674	AACGAAAAAGAGCAGTAGCTACAGATACTCGATACCTTCCCATCGCAACCCCAAAAAAGT 733
QY	721	TATCACACCGGAAAGAGTCACCGAGCGTCTACTCCAGAAAGTCAGTATGTGTGTGT 780
Db	734	TATCACACCGGAAAGAGTCACCGAGCGTCTACTCCAGAAAGTCAGTATGTGTGTGT 793
QY	781	ATGTTTTTTTAACTTTACTATAAAAGCCATGCAAAATGACAAAAATCTATATTACTTTCTCA 840
Db	794	ATGTTTTTTTAACTTTACTATAAAAGC----- 819
QY	841	AAATGGACCCCAAGAAACCTTTGATTACTGTCTTAAGTGCCTAATCTTAATTACAGGA 900
Db	820	----- 819
QY	901	ACTGTGCATCAGCTATTTATGATTCTATAAGCTATTTCAGCAGAAATGAGATATTAAACCC 960
Db	820	----- 819
QY	961	AATGCTTTGATTGTTCTAGAAAGTATAGTAATTGTTTTTCTAAGGTGGTTCAAGCATCTA 1020
Db	820	----- 819
QY	1021	CTCTTTTTTATCATTTACTTCAAAATGACATTGCTTAAAGACTGCATATTATTTTACTACTGTA 1080
Db	820	-----CATTTACTTCAAAATGACATTGCTTAAAGACTGCATATTATTTTACTACTGTA 869
QY	1081	ATTTCTCCAGCATAGCAATTATGTACATAGATGAGTGTAAACATTATATCTCACATAGA 1140
Db	870	ATTTCTCCAGCATAGCAATTATGTACATAGATGAGTGTAAACATTATATCTCACATAGA 929
QY	1141	GACATGCTTATATGGTTTTTATTTAAATGAAATGCCAGTCCATTACACTGAATAAATAGA 1200
Db	930	GACATGCTTATATGGTTTTTATTTAAATGAAATGCCAGTCCATTACACTGAATAAATAGA 989
QY	1201	ACTCAACTATTGCTTTTCAGGGAAATCATGGATAGGGTTGAAGAGGTTACTATTAAATTG 1260
Db	990	ACTCAACTATTGCTTTTCAGGGAAATCATGGATAGGGTTGAAGAGGTTACTATTAAATTG 1049
QY	1261	TTTAAAAACAGCTTAGGATTAATGTCTCTCCATTATAATGAAGATTAATAATGAAGGCTT 1320
Db	1050	TTTAAAAACAGCTTAGGATTAATGTCTCTCCATTATAATGAAGATTAATAATGAAGGCTT 1109
QY	1321	TAATCAGCATTTAAAGGAAATGAATGGCTTTCTGATATGCTGTTTTTTAGCCTAGGAG 1380
Db	1110	TAATCAGCATTTAAAGGAAATGAATGGCTTTCTGATATGCTGTTTTTTAGCCTAGGAG 1169
QY	1381	TTAGAAATCCTAACTTCTTTATCCTCTTCTCCAGAGGCTTTTTTTTCTTGTGTATTAA 1440
Db	1170	TTAGAAATCCTAACTTCTTTATCCTCTTCTCCAGAGGCTTTTTTTTCTTGTGTATTAA 1229
QY	1441	ATTAACATTTTAAACCGCAGATATTTTGTCAAGGGCTTTTGCAATCAAACTGCTTTTCC 1500
Db	1230	ATTAACATTTTAAACAGCAGATATTTTGTCAAGGGCTTTTGCAATCAAACTGCTTTTCC 1289
QY	1501	AGGCTTACTCAGAAGAAAGATAAAAGTGTGATCTAAGAAAAAGTGGTTTTTAGGAA 1560
Db	1290	AGGGCTACTCAGAAGAAAGATAAAAGTGTGATCTAAGAAAAAGTGGTTTTTAGGAA 1349
QY	1561	AGTGAATAATTTTTTGTGTTTGTATTGTAAGAAAGATGATGCATTTTGACAAGAAATCAT 1620
Db	1350	AGTGAATAATTTTTTGTGTTTGTATTGTAAGAAAGATGATGCATTTTGACAAGAAATCAT 1409
QY	1621	ATATGTATGGATATATTTTAAATAAGTATTTTGAAGTACAGACTTTGAGGTTTTCATCAATATA 1680
Db	1410	ATATGTATGGATATATTTTAAATAAGTATTTTGAAGTACAGACTTTGAGGTTTTCATCAATATA 1469

Qy	1681	AATAAAAGAGCAGAAAAATATGTCTTGGTTTTTCATTTGCTTACCAAAAAAACAACAA	1740
Db	1470	AATAAAAGAGCAGAAAAATATGTCTTGGTTTTTCATTTGCTTACCAAAAAAACAACAA	1529
Qy	1741	AAAAAGTTGTCTTTTGAGAACTTCACCTGCTCCTATGTGGGTACCTGAGTCAAAATGTG	1800
Db	1530	AAAAAGTTGTCTTTTGAGAACTTCACCTGCTCCTATGTGGGTACCTGAGTCAAAATGTG	1589
Qy	1801	ATTTTGTCTGTGAAAAATAAATTCCTCTCTGTACCATTTCTGTTTAGTTTTACTAAA	1860
Db	1590	ATTTTGTCTGTGAAAAATAAATTCCTCTCTGTACCATTTCTGTTTAGTTTTACTAAA	1649
Qy	1861	ATCTGTAAATACGTATTTTTCTGTTTATTCCAAATTTTGATGAAACTGACAATCCAA	1920
Db	1650	ATCTGTAAATACGTATTTTTCTGTTTATTCCAAATTTTGATGAAACTGACAATCCAA	1709
Qy	1921	GAAAGTTTGTGTCGACGTCGTCTAGCTTAAATGAATGTGTTCTATTGCTTTTATACAT	1980
Db	1710	GAAAGTTTGTGTCGACGTCGTCTAGCTTAAATGAATGTGTTCTATTGCTTTTATACAT	1769
Qy	1981	TATATTAATAAATTTGTACATTTTTCTAATT	2010
Db	1770	TATATTAATAAATTTGTACATTTTTCTAATT	1799

RESULT 15	AP001846	191923 bp	DNA	linear	HTG 30-MAY-2000
LOCUS	AP001846	Homo sapiens chromosome 18 clone RP11-796C24 map 18q22, WORKING DRAFT SEQUENCE, 56 unordered pieces.			
DEFINITION	AP001846	Homo sapiens chromosome 18 clone RP11-796C24 map 18q22, WORKING DRAFT SEQUENCE, 56 unordered pieces.			
ACCESSION	AP001846	GI:8117509			
VERSION	AP001846.2	HTGS_PHASE1; HTGS_DRAFT.			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 191923)				
TITLE	Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.				
JOURNAL	Homo sapiens 191,923 genomic DNA of 18q22				
REFERENCE	Published Only in DataBase (2000)				
AUTHORS	2 (bases 1 to 191923)				
TITLE	Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.				
JOURNAL	Direct Submission				
COMMENT	Submitted (18-APR-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1 Kitasato, Sagami-hara, Kanagawa 228-8555, Japan (E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923, Fax: 81-42-778-9924)				
COMMENT	On May 30, 2000 this sequence version replaced gi:7630218.				

NOTE: This is a 'working draft' sequence. It currently consists of 56 contigs. The true order of the pieces is not known and their

order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1	10920	contig	of	10920	bp	in	length
11021	18635	contig	of	7615	bp	in	length
18736	27618	contig	of	8883	bp	in	length
27719	3370	contig	of	5652	bp	in	length
33471	38522	contig	of	5052	bp	in	length
38623	44935	contig	of	6313	bp	in	length
45036	52903	contig	of	7868	bp	in	length
53004	59933	contig	of	6930	bp	in	length
60034	65715	contig	of	5682	bp	in	length
65816	73615	contig	of	7800	bp	in	length
73716	79180	contig	of	5465	bp	in	length
79281	83364	contig	of	4084	bp	in	length
83465	874054	contig	of	3590	bp	in	length
87155	91199	contig	of	4045	bp	in	length
91300	94966	contig	of	3667	bp	in	length
95067	99407	contig	of	4341	bp	in	length
99508	103203	contig	of	3696	bp	in	length
103304	106888	contig	of	3585	bp	in	length
106989	110187	contig	of	3199	bp	in	length
110288	113798	contig	of	3511	bp	in	length
113899	117663	contig	of	3765	bp	in	length
117764	121135	contig	of	3372	bp	in	length
121236	123224	contig	of	1989	bp	in	length
123325	125604	contig	of	2280	bp	in	length
125705	127990	contig	of	2286	bp	in	length
128091	130683	contig	of	2593	bp	in	length
130784	133371	contig	of	2588	bp	in	length
133472	136426	contig	of	2955	bp	in	length
136527	139256	contig	of	2730	bp	in	length
139357	141388	contig	of	2032	bp	in	length
141489	144089	contig	of	2601	bp	in	length
144190	146686	contig	of	2497	bp	in	length
146787	149873	contig	of	3087	bp	in	length
149974	151895	contig	of	1922	bp	in	length
151996	154840	contig	of	2845	bp	in	length
154941	157109	contig	of	2169	bp	in	length
157210	159650	contig	of	2441	bp	in	length
159751	161167	contig	of	1417	bp	in	length
161268	163443	contig	of	2176	bp	in	length
163544	165635	contig	of	2092	bp	in	length
165736	168432	contig	of	2697	bp	in	length
168533	171331	contig	of	1584	bp	in	length
171432	173015	contig	of	1728	bp	in	length
173116	174843	contig	of	1342	bp	in	length
174944	176285	contig	of	1243	bp	in	length
176386	177628	contig	of	1694	bp	in	length
177329	179422	contig	of	1184	bp	in	length
179523	180706	contig	of	1409	bp	in	length
180807	182215	contig	of	1358	bp	in	length
182316	183673	contig	of	1260	bp	in	length
183774	185033	contig	of	1768	bp	in	length
185134	186901	contig	of	1366	bp	in	length
187002	188367	contig	of	1453	bp	in	length
189468	189920	contig	of	1312	bp	in	length
190021	191332	contig	of	491	bp	in	length
191433	191923	contig	of				

Sequence updated (26-May-2000).

```

sequence updated as they are added.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

* 1 10920: contig of 10920 bp in length
 * 10921 11020: gap of 100 bp
 * 11021 18635: contig of 7615 bp in length

Qy	901	ACTGTGCATCAGCTATTTATGATTTCTATAAGCTATTTTACAGCAGAAATGAGATATATAAACCC	960
Db	130213	ACTGTGCATCAGCTATTTATGATTTCTATAAGCTATTTTACAGCAGAAATGAGATATATAAACCC	130272
Qy	961	AATGCTTTGATTGTTCTAGAAAGTATAGTAATTTTGTTTCTAAAGGTGGTTCAAGCATCTA	1020
Db	130273	AATGCTTTGATTGTTCTAGAAAGTATAGTAATTTTGTTTCTAAAGGTGGTTCAAGCATCTA	130332
Qy	1021	CTCTTTTTATCAATTTACTTCAAAAATGACATTTGCTAAAGACTGCATTTATTTTACTACTGTA	1080
Db	130333	CTCTTTTTATCAATTTACTTCAAAAATGACATTTGCTAAAGACTGCATTTATTTTACTACTGTA	130392
Qy	1081	ATTTCTCCACGACATAGCATTATGTACATAGATGAGTGTAAACATTTATATCTCACATAGA	1140
Db	130393	ATTTCTCCACGACATAGCATTATGTACATAGATGAGTGTAAACATTTATATCTCACATAGA	130452
Qy	1141	GACATGCTTATATGGTTTTTATTTAAATGAAATGCCAGTCCATTACACTGAATAAAATAGA	1200
Db	130453	GACATGCTTATATGGTTTTTATTTAAATGAAATGCCAGTCCATTACACTGAATAAAATAGA	130512
Qy	1201	ACTCAACTATGCTTTTTCAGGGAAATCATGGATAGGGTTGAAGAAGGTTACTATTAAATTG	1260
Db	130513	ACTCAACTATGCTTTTTCAGGGAAATCATGGATAGGGTTGAAGAAGGTTACTATTAAATTG	130572
Qy	1261	TTTAAAAACAGCTTAGGGATTAAATGTCCTCCATTATATGAAGATTAAATGAAGGCTT	1320
Db	130573	TTTAAAAACAGCTTAGGGATTAAATGTCCTCCATTATATGAAGATTAAATGAAGGCTT	130632
Qy	1321	TAATCAGCATTTGTAAAGGAAATTGAATGGCTTTCTGATATGCTGTTTT	1368
Db	130633	TAATCAGCATTTGTAAAGGAAATTGAATGGCTTTCTGATATGCTGTTTT	130680

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2004, 22:04:00 ; Search time 3550 Seconds
(without alignments)

16907.883 Million cell updates/sec

Title: US-10-063-732-119

Perfect score: 2010
Sequence: 1 ggaataactgtctctctctg.....aattgtacatttttctaatt 2010

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_mam:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_rod:*
- 26: em_gss_phg:*
- 27: em_gss_vrl:*
- 28: gb_gss1:*
- 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	761	37.9	837	9 AU121779	AU121779 AU121779
2	719	35.8	754	12 BG674625	BG674625 602620674
3	716	35.6	772	12 BG426438	BG426438 602492836
C 4	683.4	34.0	703	14 CB851106	CB851106 UI-CF-EN1

5	678	33.7	29	AY400665	AY400665 Homo sapi
6	587.6	29.2	12	BG402103	BG402103 602465638
C 7	552	27.5	10	AW235670	AW235670 xn21c12.x
C 8	542.4	27.0	10	BF195920	BF195920 7087c12.x
C 9	541	26.9	10	BF196092	BF196092 hr81h10.x
C 10	535.8	26.7	9	AU147533	AU147533 AU147533
11	525.2	26.1	14	CB851224	CB851224 UI-CF-DU1
C 12	514.4	25.6	12	BG957666	BG957666 PMO-CT080
C 13	512	25.5	10	AW237740	AW237740 xm81a03.x
14	505.2	25.1	10	BF675808	BF675808 602083725
15	494	24.6	29	AY400667	AY400667 Mus muscu
C 16	480.4	23.9	492	AW205135	AW205135 UI-H-B11-
17	468.4	23.3	14	CB418866	CB418866 591692 MA
18	462.2	23.0	14	CF766535	CF766535 CES003314
19	460.2	22.9	14	CF169716	CF169716 B0817B10-
20	445.2	22.1	12	BG402412	BG402412 602466111
21	443.2	22.0	12	BI852250	BI852250 603378454
22	440.2	21.9	557	BF231401	BF231401 253787 BA
23	438.8	21.8	14	CA313163	CA313163 UI-CF-FN0
C 24	429	21.3	14	CF796030	CF796030 892243 MA
25	427.8	21.3	14	CF169015	CF169015 B0808A07-
C 26	425	21.1	9	AA536178	AA536178 nf96b03.s
27	421.6	21.0	14	CF140548	CF140548 UI-HF-CB0
28	404.4	20.1	752	AJ507044	AJ507044 AJ507044
C 29	399	19.9	10	AW237781	AW237781 xm81e07.x
30	398.6	19.8	14	CK333041	CK333041 H8229C06-
31	397.4	19.8	14	CF168607	CF168607 B0802A03-
32	396.6	19.7	551	AI786907	AI786907 uj32g11.y
33	390.8	19.4	12	BI104115	BI104115 602890432
34	365.8	18.2	14	CB455630	CB455630 712787 MA
35	360.8	18.0	14	CB425534	CB425534 600565 MA
36	360.4	17.9	10	BE285566	BE285566 601095576
37	355.6	17.7	14	CB235084	CB235084 AGENCOURT
38	355	17.7	13	BX283446	BX283446 BX283446
39	351	17.5	360	BE172635	BE172635 MR0-HT055
40	347.2	17.3	352	BE000282	BE000282 MR0-BN007
C 41	345.2	17.2	351	BE814313	BE814313 MR0-BN007
C 42	339.8	16.9	10	BB618773	BB618773 BB618773
43	336	16.7	363	CB306981	CB306981 UI-CF-FN0
44	332.4	16.5	568	CA339169	CA339169 NISC_1x11
45	330.6	16.4	465	AI048929	AI048929 uc84g08.y
46	328.2	16.3	771	AI547146	AI547146 PN3.1_01
47	325.2	16.2	643	BB610589	BB610589 BB610589
48	323.4	16.1	325	AY400666	AY400666 Pan trogl
49	319.8	15.9	578	BF151294	BF151294 uz14e04.y
C 50	311.2	15.5	348	AA769736	AA769736 ob20c04.s
51	311.2	15.5	375	CF140477	CF140477 UI-HF-CB0
C 52	284.4	14.1	459	BI403157	BI403157 MI-P-CF1-
53	284.2	14.1	626	CF171768	CF171768 B0847B12-
54	282.6	14.1	500	BE483534	BE483534 169725 BA
55	278.8	13.9	862	EX720601	EX720601 BX720601
56	278.2	13.8	482	BQ265325	BQ265325 NISC_1f05
57	277.6	13.8	675	AY401721	AY401721 Homo sapi
58	276.4	13.8	613	CF171518	CF171518 B0843E03-
59	271.2	13.5	675	AY401722	AY401722 Pan trogl
60	270.6	13.5	1174	AK048287	AK048287 Mus muscu
61	269.4	13.4	856	CB236303	CB236303 AGENCOURT
62	267.8	13.3	675	AY401723	AY401723 Mus muscu
63	256.8	12.8	268	AW999412	AW999412 MR0-BN007
C 64	251.2	12.5	256	BE172270	BE172270 MR0-HT055
65	245.2	12.2	465	AI037192	AI037192 ud60a11.y
C 66	243.8	12.1	642	CB457008	CB457008 714331 MA
C 67	241.6	12.0	648	BZ846477	BZ846477 CH240_193
68	239.4	11.9	316	BB476649	BB476649 160002 BA
C 69	233.8	11.6	266	BF329750	BF329750 RC0-BN028
C 70	233.6	11.6	593	CB433030	CB433030 609301 MA
71	233.2	11.6	441	AA839199	AA839199 vw93h03.r
C 72	225	11.2	682	BI401450	BI401450 MI-P-CP0-
73	221.2	11.0	397	AI640028	AI640028 ms65h08.y
74	221	11.0	521	AW990108	AW990108 uf34c01.y
75	219.8	10.9	402	AA167924	AA167924 ms65h08.r
76	217	10.8	585	CF114057	CF114057 Shultzomi
77	215.4	10.7	535	AW226871	AW226871 um62g12.y

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10594 row: c column: 19
High quality sequence stop: 750.
Location/Qualifiers

1. .754

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4746186"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Skn3"
/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI CGAP Library."

ORIGIN

Query Match 35.8%; Score 719; DB 12; Length 754;
Best Local Similarity 99.6%; Pred. No. 1e-133;
Matches 752; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

QY 471 CTTTCATCATCAGGGGATGGTGGTCTCATCCCTGTGAGCTGGTGGCAATGCCATCAT 530
Db 1 CTTTCATCATCAGGGGATGGTGGTCTCATCCCTGTGAGCTGGTGGCAATGCCATCAT 60
QY 531 CAGAGATTTCTATACTCAATAGTAGTAATGTGCCCAAAACGTGAGCTTGGAGAAGCTCT 590
Db 61 CAGAGATTTCTATACTCAATAGTAGTAATGTGCCCAAAACGTGAGCTTGGAGAAGCTCT 120
QY 591 CTACTTAGGATGGACCCAGGCACCTGGTCTGATTTGGAGAGCTCTG-TTCTGCTGCG 649
Db 121 CTACTTAGGATGGACCCAGGCACCTGGTCTGATTTGGAGAGCTCTGTTTCTGCTGCG 180
QY 650 TTTTGTGTGCAACGAAAGAGCAGTAGCTACAGATACCTCGATACCTTCCCATCGCACAA 709
Db 181 TTTTGTGTGCAACGAAAGAGCAGTAGCTACAGATACCTCGATACCTTCCCATCGCACAA 240
QY 710 CCCAAAAAGTTATCACACCGGAAAGAGTCCAGCGCTCTACTCCAGAAGTCAGTATG 769
Db 241 CCCAAAAAGTTATCACACCGGAAAGAGTCCAGCGCTCTACTCCAGAAGTCAGTATG 300
QY 770 TGTAGTTGTGTATGTTTTTTTAACTTTACTATAAAGCCATGCAAAATGACAAAAATCTATA 829
Db 301 TGTAGTTGTGTATGTTTTTTTAACTTTACTATAAAGCCATGCAAAATGACAAAAATCTATA 360
QY 830 TTACTTTCTCAAAATGGACCCCAAGAACTTTGATTACTGTGTTTAACTGCCCTAATCT 889
Db 361 TTACTTTCTCAAAATGGACCCCAAGAACTTTGATTACTGTGTTTAACTGCCCTAATCT 420
QY 890 TAAATACAGGAACCTGTGCATCAGCTATTTATGATTTCTATAAGCTATTTTCAGCAGATGAG 949
Db 421 TAAATACAGGAACCTGTGCATCAGCTATTTATGATTTCTATAAGCTATTTTCAGCAGATGAG 480
QY 950 ATATTAAACCCCAATGCTTTGATTTGTTCTAGAAAGTATAGTAATTTGTTTCTAAGTGGT 1009
Db 481 ATATTAAACCCCAATGCTTTGATTTGTTCTAGAAAGTATAGTAATTTGTTTCTAAGTGGT 540
QY 1010 TCAAGCATCTACTCTTTTATCATTTACTTCAAAATGACATTCGCTAAAGACTGCATTATT 1069
Db 541 TCAAGCATCTACTCTTTTATCATTTACTTCAAAATGACATTCGCTAAAGACTGCATTATT 600
QY 1070 TTACTACTGTAATTTCTCCAGCATAGCATATATGTACATAGATGAGTGT-AACATTTAT 1128
Db 601 TTACTACTGTAATTTCTCCAGCATAGCATATATGTACATAGATGAGTGTAAACATTTAT 660
QY 1129 ATCTCACATAGACATGCTTATATGTTTATTTTAAATGAAATGCCAGTCCATTACAC 1188

Db 661 ATCTCACATAGACATGCTTATATGTTTATTAATAATGAAATGCCAGTCCATTACAC 720
QY 1189 TGAATAAATAGAACTCAACTATTGCTTTTCAGGGA 1223
Db 721 TGAATAAATAGAACTCAA-TATTGCTTTTCAGGGA 754

RESULT 3
BG426438

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BG426438 772 bp mRNA linear EST 14-MAR-2001
602492836F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4606880 5',
mRNA sequence.

BG426438 GI:13332944

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 772)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM1346 row: g column: 09

High quality sequence stop: 735.

Location/Qualifiers

FEATURES

source

1. .772

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4606880"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NIH_MGC_75"

/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggccattatggcc); Site 2: SfiI (ggccattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGCGGCGGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.65
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 35.6%; Score 716; DB 12; Length 772;
Best Local Similarity 99.3%; Pred. No. 4.1e-133;
Matches 760; Conservative 0; Mismatches 1; Indels 4; Gaps 4;

QY 25 ACAGAGAACCCCTGCTTCAAAGCAGAAAGTAGCAGTTCGGAGTCCAGCTGGCTAAACTCA 84
Db 1 ACAGAGAACCCCTGCTTCAAAGCAGAAAGTAGCAGTTCGGAGTCCAGCTGGCTAAACTCA 60
QY 85 TCCCAGAGGATAATGGCAACCCATGCCCTTAGAAATCGCTGGGCTGTTTCTTGGTGGTGT 144
Db 61 TCCCAGAGGATAATGGCAACCCATGCCCTTAGAAATCGCTGGGCTGTTTCTTGGTGGTGT 120
QY 145 GGAATGGTGGGACAGTGGCTGTCACTGTCTAGTCCCTCAGTGGAGAGTGTGGCCTTCATT 204
Db 121 GGAATGGTGGGACAGTGGCTGTCACTGTCTAGTCCCTCAGTGGAGAGTGTGGCCTTCATT 180
QY 205 GAAACCAACATCGTGGTGTGTTTGAATACTTCTGGGAAGGACTGTGGATGAATTCGTGAGG 264

Db 181 GAAACAACATCGTGGTTTTTGAACACTTCTGGGAGGACTGTGGATGAATTCGTGAGG 240

QY 265 CAGGCTAACATCAGGATGCAGTGAACAAATCTATGATTCCTGCTGGCTCTTCTCCGGAC 324

Db 241 CAGGCTAACATCAGGATGCAGTGAACAAATCTATGATTCCTGCTGGCTCTTCTCCGGAC 300

QY 325 CTACAGGCAGCCAGGACTGATGTGTGCTGCTTCCTGATGTCCTTCTTGGCTTTCATG 384

Db 301 CTACAGGCAGCCAGGACTGATGTGTGCTGCTTCCTGATGTCCTTCTTGGCTTTCATG 360

QY 385 ATGGCCATCCTTGGCATGAAATGCACACAGGTGCACGGGGACAATGAGAAGGTGAAGGCT 444

Db 361 ATGGCCATCCTTGGCATGAAATGCACACAGGTGCACGGGGACAATGAGAAGGTGAAGGCT 420

QY 445 CACATTCTGCTGACGGCTGGAATCATCTTCATCATCAGGGGCATGGTGGTCTCATCCCT 504

Db 421 CACATTCTGCTGACGG-TGGAATCATCTTCATCATCAGGGGCATGGTGGTCTCATCCCT 479

QY 505 GTGAGCTGGTGGCCAAATGCCATCATCAGAGATTTCTATAACTCAATAGTGAATGTTGCC 564

Db 480 GTGAGCTGGTGGCCAAATGCCATCATCAGAGATTTCTATAACTCAATAGTGAATGTTGCC 539

QY 565 CAAAAACGTGAGCTTGGAGAAGTCTCTACTTAGATGGACCCAGGACTGGTGTGCTGATT 624

Db 540 CAAAAACGTGAGCTTGGAGAAGTCTCTACTTAGATGGACCCAGGACTGGTGTGCTGATT 599

QY 625 GTTGAGGAGCTCTGTTCTGCTGCGTTTTTTGTTGCAACGAAAGACAGTAGCTACAGA 684

Db 600 GTTGAGGAGCTCTGTTCTGCTGCG-TTTTTGTGCAACGAAAGACAGTAGCTACAGA 658

QY 685 TACTCGATACCTTCCCATCGCAACAAACCCAAAAGTTATCACACCGGAAAGAGTCACCG 744

Db 659 TACTCGATACCTTCCCATCG-ACAANCCAAAAGTTATCACACCGGAAAGAGTCACCG 717

QY 745 AGCGTCTACTCCAGAAAGTC-AGPATGTGTAGTTGTGTATGTTTT 788

Db 718 AGCGTCTACTCCAGAAAGTCAAGTATGTGTAGTTGTGTATGTTTT 762

RESULT 4

CB851106/c

LOCUS

DEFINITION UI-CF-EN1-adg-i-18-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone

UI-CF-EN1-adg-i-18-0-UI 3', mRNA sequence.

ACCESSION CB851106

VERSION CB851106.1 GI:30045875

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 703)

TITLE Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

PUBMED 8889548

COMMENT Contact: McCray, PB

McCray Lab

University of Iowa

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).

The following repetitive elements were found in this cDNA sequence: 1-58, >At_rich#low_complexity (matched complement)

FEATURES

source

Location/Qualifiers

1..703

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-CF-EN1-adg-i-18-0-UI"

/tissue_type="Primary Lung Cystic Fibrosis Epithelial Cells"

/dev_stage="Adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone_lib="UI-CF-EN1"

/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-CF-EN1 is a normalized cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGGT.

TAG_SEQ=None found"

ORIGIN

Query Match 34.0%; Score 683.4; DB 14; Length 703;

Best Local Similarity 99.9%; Pred. No. 1.4e-126;

Matches 684; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1326 AGCATTGTAAAGGAAATGAATGGCTTCTGATATGCTGTTTTTAGCTAGGAGTTAGA 1385

Db 703 AGCATTGTAAAGGAAATGAATGGCTTCTGATATGCTGTTTTTAGCTAGGAGTTAGA 644

QY 1386 AATCCTAACTTCTTATCCTCTTCTCCAGAGGCTTTTTTTTCTTGTGATTAATAA 1445

Db 643 AATCCTAACTTCTTATCCTCTTCTCCAGAGGCTTTTTTTTCTTGTGATTAATAA 584

QY 1446 CATTTTAAACGCAGATATTTTGTCAAGGGCTTTCATTCAAACGCTGTTTCCAGGGC 1505

Db 583 CATTTTAAACGCAGATATTTTGTCAAGGGCTTTCATTCAAACGCTGTTTCCAGGGC 524

QY 1506 TATACTCAGAGAAAGATATAAAGTGTGATCTAAGAAAAAGTGATGGTTTAGGAAAGTGA 1565

Db 523 TATACTCAGAGAAAGATATAAAGTGTGATCTAAGAAAAAGTGATGGTTTAGGAAAGTGA 464

QY 1566 AAATATTTTGTGTTTGTATTTGAAGAAGATGATGTCATTTTGACAAGAAATCATATATG 1625

Db 463 AAATATTTTGTGTTTGTATTTGAAGAAGATGATGTCATTTTGACAAGAAATCATATATG 404

QY 1626 TATGGATATATTTAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATATAAATAA 1685

Db 403 TATGGATATATTTAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATATAAATAA 344

QY 1686 AAGAGCAGAAAAAATATGCTCTGGTTTTTCATTTGCTTACCAAAAAACAACAAAAAAA 1745

Db 343 AAGAGCAGAAAAAATATGCTCTGGTTTTTCATTTGCTTACCAAAAAACAACAAAAAAA 284

QY 1746 GTTGTCTTTCAGAACTTCACCTGCTCCTATGTGGGTACCTGAGTCAAAATTTGTCATTTT 1805

Db 283 GTTGTCTTTCAGAACTTCACCTGCTCCTATGTGGGTACCTGAGTCAAAATTTGTCATTTT 224

QY 1806 TGTTCTGTGAAAAATAAATTTCTTGTACCATTTCTGTTAGTTTACTAAATCTG 1865

Db 223 TGTTCTGTGAAAAATAAATTTCTTGTACCATTTCTGTTTACTAAATCTG 164

QY 1866 TAAATACGTATTTTTCTGTTTATTTCCAAATTTGATGAAACTGACAAATTTGAAAG 1925

Db 163 TAAATACGTATTTTTCTGTTTATTTCCAAATTTGATGAAACTGACAAATTTGAAAG 104

7

```

RESULT 6
BG402103
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
    source
        BG402103
        602465638F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4593920 5',
        mRNA sequence.
        BG402103
        BG402103.1 GI:13295551
        EST.
        Homo sapiens (human)
        Homo sapiens
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
        1 (bases 1 to 903)
        NIH-MGC http://mgc.nci.nih.gov/.
        National Institutes of Health, Mammalian Gene Collection (MGC)
        Unpublished (1999)
        Contact: Robert Strausberg, Ph.D.
        Email: cgabbs@email.nih.gov
        Tissue Procurement: CLONTECH Laboratories, Inc.
        cDNA Library Preparation: CLONTECH Laboratories, Inc.
        cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
        DNA Sequencing by: Incyte Genomics, Inc.
        Clone distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LLNL at:
        http://image.llnl.gov
        Plate: LLC1334 row: k column: 09
        High quality sequence stop: 599.
        Location/Qualifiers
            1..903
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone=IMAGE:4593920"
            /lab_host="DH10B (TI phage-resistant)"
            /clone.lib="NIH_MGC_75"
            /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
            sfii (ggcgccctcgcc); Site 2: sfii (ggccattatggcc); 5' and
            3' adaptors were used in cloning as follows: 5' adaptor
            sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
            5'-ATTCTAGAGCCGAGGCGCGACATG-dt(30)BN-3' (where B = A,
            C, or G and N = A, C, G, or T). Average insert size 1.65
            kb (range 0.5-4.0 kb). 15/15 colonies contained inserts

```


by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

ORIGIN

Query Match	29.2%;	Score 587.6;	DB 12;	Length 903;
Best Local Similarity	99.0%;	Pred. No. 2.1e-107;		
Matches 612;	Conservative 0;	Mismatches 4;	Indels 2;	Gaps 2;
Qy	242	GACTGTGGATGAATTGCGTGAGGAGGCTAACATCAGGATGCAGTGCAAAATCTATGATT	301	
Db	1	GACTGTGGATGAATTGCGTGAGGAGGCTAACATCAGGATGCAGTGCAAAATCTATGATT	60	
Qy	302	CCCTGCTGGCTCTTTCTCCGGACCTACAGGCAGCCAGGACTGATGTGCTGCTTCCG	361	
Db	61	CCCTGCTGGCTCTTTCTCCGGACCTACAGGCAGCCAGGACTGATGTGCTGCTTCCG	120	
Qy	362	TGATGTCCTTCTTGGCTTTCATGATGGCCATCCTTGGCATGAAATGCACGAGTGCACGG	421	
Db	121	TGATGTCCTTCTTGGCTTTCATGATGGCCATCCTTGGCATGAAATGCACGAGTGCACGG	180	
Qy	422	GGGACAAATGAGAAGGTGAAGGCTCACATTCTGCTGACGGCTGGAATCATCTTCATCATCA	481	
Db	181	GGGACAAATGAGAAGGTGAAGGCTCACATTCTGCTGACGGCTGGAATCATCTTCATCATCA	240	
Qy	482	CGGGCATGGTGGTCTCATCCCTGTGAGCTGGTGGCAATGCCATCATCAGAGATTCT	541	
Db	241	CGGGCATGGTGGTCTCATCCCTGTGAGCTGGTGGCAATGCCATCATCAGAGATTCT	300	
Qy	542	ATAACTCAATAGTGAATGTTGCCCAAAACGTGAGCTGGAGAGCTCTCTACTTAGGAT	601	
Db	301	ATAACTCAATAGTGAATGTTGCCCAAAACGTGAGCTGGAGAGCTCTCTACTTAGGAT	360	
Qy	602	GGACCACGGCACTGGTGTGCTGATTGTTGGAGGAGCTCTGTTCTGCTGGCTTTTGTGCA	661	
Db	361	GGACCACGGCACTGGTGTGCTGATTGTTGGAGGAGCTCTGTTCTGCTGGCTTTTGTGCA	420	
Qy	662	ACGAAAAGAGCAGTAGCTACAGATCTCGATACCTTCCCATCGCACACCCCAAAAAGTT	721	
Db	421	ACGAAAAGAGCAGTAGCTACAGATCTCGATACCTTCCCATCGCACACCCCAAAAAGTT	480	
Qy	722	ATCACACCGGAAAGAGTCACCGAGCGTCTACTCCAGAAAGTCAGTATGTGTGTA	781	
Db	481	ATCACACCGGAAAGAGTCACCGAGCGTCTACTCCAGAAAGTCAGTATGTGTGTA	540	
Qy	782	TGTTTTTTAACTTTACTATAAAGCCATGCAAAATGACAAAAATCTATATTAC-TTTCTCA	840	
Db	541	TG-TTTTTTAACTTTACTATAAAGCCATGCAAAATGACAAAAATCTATATTACTTTTCTCA	599	
Qy	841	AAATGGACCCCAAGAAA 858		
Db	600	AAATGGACCCCAAGAAA 617		

RESULT 7
AW235670/c
LOCUS
DEFINITION xn21c12.x1 NCI_CGAP_Kid11 Homo sapiens cdna clone EST 13-DEC-1999
mRNA sequence.
AW235670
VERSION AW235670.1 GI:6568059
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 559)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 444.

FEATURES

Source

Location/Qualifiers
1..559
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2694358"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Kid11"
/note="Organ: Kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match	27.5%;	Score 552;	DB 10;	Length 559;
Best Local Similarity	100.0%;	Pred. No. 3.1e-100;		
Matches 552;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1459	CAGATATTTTGTCAAGGGGCTTTGCAATTCAAACTGCTTTCCAGGGCTACTCAGAAGA	1518	
Db	559	CAGATATTTTGTCAAGGGGCTTTGCAATTCAAACTGCTTTCCAGGGCTACTCAGAAGA	500	
Qy	1519	AAGATAAAAGTGTGATCTAAGAAAAGTGATGGTTTTAGGAAAGTGAAAATATTTTGT	1578	
Db	499	AAGATAAAAGTGTGATCTAAGAAAAGTGATGGTTTTAGGAAAGTGAAAATATTTTGT	440	
Qy	1579	TTTGTATTTGAAGAAGATGATGATTTTGACAAGAAATCATATATGTATGGATATATT	1638	
Db	439	TTTGTATTTGAAGAAGATGATGATTTTGACAAGAAATCATATATGTATGGATATATT	380	
Qy	1639	TAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATATAATAAAGAGCAGAAAA	1698	
Db	379	TAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATATAATAAAGAGCAGAAAA	320	
Qy	1699	TATGCTCTGGTTTTTCATTTGCTTACCAAAAAACAACAAGAGTTGCTCTTGAG	1758	
Db	319	TATGCTCTGGTTTTTCATTTGCTTACCAAAAAACAACAAGAGTTGCTCTTGAG	260	
Qy	1759	AACCTCACCTGCTCTATGTGGGTACCTGAGTCAAAATGTCATTTTGTCTGTGAAAA	1818	
Db	259	AACCTCACCTGCTCTATGTGGGTACCTGAGTCAAAATGTCATTTTGTCTGTGAAAA	200	
Qy	1819	ATAAATTTCTCTTGTACCATTTCTGTTTAGTTTACTAAAACTCTGTAATACTGTATT	1878	
Db	199	ATAAATTTCTCTTGTACCATTTCTGTTTAGTTTACTAAAACTCTGTAATACTGTATT	140	
Qy	1879	TTTCTGTTTATCCAAATTTGATGAAACTGACAAATCCAAATTTGAAAGTTGTGCGCGT	1938	
Db	139	TTTCTGTTTATCCAAATTTGATGAAACTGACAAATCCAAATTTGAAAGTTGTGCGCGT	80	
Qy	1939	CTGTCTAGCTTAAATGAATGTGTTCTATTGCTTTTACATTTATTAATAAATTTGTAC	1998	
Db	79	CTGTCTAGCTTAAATGAATGTGTTCTATTGCTTTTACATTTATTAATAAATTTGTAC	20	
Qy	1999	ATTTTCTTAATT 2010		
Db	19	ATTTTCTTAATT 8		

RESULT 8
BF195920/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
BF195920 544 bp mRNA linear EST 03-NOV-2000
7087c12.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3643390 3',
mRNA sequence.
BF195920
BF195920.1 GI:11083309
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 544)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
High quality sequence stop: 467.
FEATURES
source
1. .544
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3643390"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Kid11"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo. "
ORIGIN
Query Match 27.0%; Score 542.4; DB 10; Length 544;
Best Local Similarity 99.8%; Pred. No. 2.6e-98;
Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1064 ATTATTTTACTACTGTAATTTCTCCACGACATAGCATTATGTACATAGATGAGTGTAAACA 1123
Db |||||||
544 ATTATTTTACTACTGTAATTTCTCCACGACATAGCATTATGTACATAGATGAGTGTAAACA 485
QY |||||||
1124 TTTATATCTCACATAGACATGCTTATATGGTTTATTTAAATGAAATGCCAGTCCAT 1183
Db |||||||
484 TTTATATCTCACATAGACATGCTTATATGGTTTATTTAAATGAAATGCCAGTCCAT 425
QY |||||||
1184 TACACTGAATAAATAGAACTCAACTATTGCTTTTCAGGGAAATCATGGATAGGGTTGAAG 1243
Db |||||||
424 TACACTGAATAAATAGAACTCAACTATTGCTTTTCAGGGAAATCATGGATAGGGTTGAAG 365
QY |||||||
1244 AAGGTTACTATTAAATGTTTTAAAAACAGCTTAGGATTAATGTCTCCATTATATGAATGAA 1303
Db |||||||
364 AAGGTTACTATTAAATGTTTTAAAAACAGCTTAGGATTAATGTCTCCATTATATGAATGAA 305
QY |||||||
1304 GATTAAATGAAGGCTTTAATCAGCATTGTAAGGAAATTTGAATGCGTTTCTGATATGCT 1363
Db |||||||
304 GATTAAATGAAGGCTTTAATCAGCATTGTAAGGAAATTTGAATGCGTTTCTGATATGCT 245

QY 1364 GTTTTTTAGCTAGGAGTTAGAAATCCTAACTTCTTTATCCTCTTCTCCAGAGGCTTTT 1423
Db |||||||
244 GTTTTTTAGCTAGGAGTTAGAAATCCTAACTTCTTTATCCTCTTCTCCAGAGGCTTTT 185
QY 1424 TTTTCTTGTGTATTAAATTAACATTTTAAACGCAGATATTTTGTCAAGGGGCTTTGC 1483
Db |||||||
184 TTTTCTTGTGTATTAAATTAACATTTTAAAGCAGATATTTTGTCAAGGGGCTTTGC 125
QY 1484 ATTCAAACTGCTTTTCCAGGGCTATACCTCAGAAGAAAGATAAAAGTGATCTAAGAAA 1543
Db |||||||
124 ATTCAAACTGCTTTTCCAGGGCTATACCTCAGAAGAAAGATAAAAGTGATCTAAGAAA 65
QY 1544 AGTGATGGTTTATAGGAAAGTGAATAATTTTGTGTTTGTATTGTAAGAGAATGATGCA 1603
Db |||||||
64 AGTGATGGTTTATAGGAAAGTGAATAATTTTGTGTTTGTATTGTAAGAGAATGATGCA 5
QY 1604 TTTT 1607
Db |||||
4 TTTT 1
RESULT 9
BF196092/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
BF196092 547 bp mRNA linear EST 03-NOV-2000
hr81h10.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3134947 3',
mRNA sequence.
BF196092
BF196092.1 GI:11083657
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 547)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
High quality sequence stop: 472.
FEATURES
source
1. .547
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3134947"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Kid11"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo. "
ORIGIN
Query Match 26.9%; Score 541; DB 10; Length 547;
Best Local Similarity 100.0%; Pred. No. 4.9e-98;
Matches 541; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1470 TCAAGGGGCTTTGCAATTCAAACTGCTTTCCAGGGCTATACCTCAGAGAAAGATAAAAGT 1529
Db |||||||
547 TCAAGGGGCTTTGCAATTCAAACTGCTTTCCAGGGCTATACCTCAGAGAAAGATAAAAGT 488
QY 1530 GTGATCTAAGAAAAAGTGATGGTTTGTAGGAAAGTGAAAAATATTTTGTGTTTGTATTGA 1589
Db |||||||
487 GTGATCTAAGAAAAAGTGATGGTTTGTAGGAAAGTGAAAAATATTTTGTGTTTGTATTGA 428
QY 1590 AGAAGAATGATGATGATTTTGACAAGAAATCATATATGATGATGATATATTTTAATAAGTATT 1649
Db |||||||
427 AGAAGAATGATGATGATTTTGACAAGAAATCATATATGATGATGATATATTTTAATAAGTATT 368
QY 1650 TGAGTACAGACTTTGAGGTTTTCATCAATATAATAAAGAGAGCAGAAAAATATGCTTTGGT 1709
Db |||||||
367 TGAGTACAGACTTTGAGGTTTTCATCAATATAATAAAGAGAGCAGAAAAATATGCTTTGGT 308
QY 1710 TTTTCATTGCTTACCACAAAAACACAAACAAAAAGTTGCTCTTTTGAGAACTTCACCTG 1769
Db |||||||
307 TTTTCATTGCTTACCACAAAAACACAAACAAAAAGTTGCTCTTTTGAGAACTTCACCTG 248
QY 1770 CTCCTATGTGGGTACCTGAGTCAGAAATGTCATTTTGTCTGTGGAATAAATTTTCCT 1829
Db |||||||
247 CTCCTATGTGGGTACCTGAGTCAGAAATGTCATTTTGTCTGTGGAATAAATTTTCCT 188
QY 1830 TCTTGTAACCATTTCTGTTTGTAGTTTACTAAATCTGTAAATCTGTATTTTCTGTTAT 1889
Db |||||||
187 TCTTGTAACCATTTCTGTTTGTAGTTTACTAAATCTGTAAATCTGTATTTTCTGTTAT 128
QY 1890 TCCAAATTTGATGAAACTGACAAATCCAAATTTGAAAGTTGTGTCGACGCTGTCTAGCTT 1949
Db |||||||
127 TCCAAATTTGATGAAACTGACAAATCCAAATTTGAAAGTTGTGTCGACGCTGTCTAGCTT 68
QY 1950 AAATGAATGTGTTCTATTGCTTTATACATTTATATAAATAAATGTCATTTTCTAAT 2009
Db |||||||
67 AAATGAATGTGTTCTATTGCTTTATACATTTATATAAATAAATGTCATTTTCTAAT 8
QY 2010 T 2010
Db 7 T 7

RESULT 10
AU147533/c
LOCUS AU147533 MAMMA1 Homo sapiens cDNA clone MAMMA1000956 3', mRNA
DEFINITION AU147533.1 GI:11009054
sequence.
ACCESSION AU147533
VERSION AU147533.1 GI:11009054
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 577)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura,Y., Nagai,T.,
Sugano,S. and Isogai,T.).
HRI human cDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S.,
Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Ozawa,M.,
Nakamura,Y., Nagai,T., Sugano,S., Isogai,T.)
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES
source
Location/Qualifiers
1..577
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MAMMA1000956"
/tissue_type="mammary gland"
/clone_lib="MAMMA1"
/note="Vector: pME18SFL3"
ORIGIN
Query Match 26.7%; Score 535.8; DB 9; Length 577;
Best Local Similarity 97.2%; Pred. No. 5.3e-97;
Matches 562; Conservative 0; Mismatches 14; Indels 2; Gaps 2;
QY 1265 AAAACAGCTTAGGGATTAAATGCTCTCCATTATATAATGAAGATTAAAAATGAAGGCTTTAAT 1324
Db |||||||
577 AAAANCAGTTAGGNATTAAAGGTCNCCTATTNATAATGANGATTAAATGAAGNCTTNATC 518
QY 1325 CAGCATTGTAAAGGAAATTTGAATGGCTTTCTGATATGCTGTTTTTTTAGCCCTAGGAGTTAG 1384
Db |||||||
517 CAGCATTG-AAAGGAAATTTGAATGGCTTTCTGATATGCTGTTTTTTTAGCCCTAGGAGTTAG 459
QY 1385 AAATCCTAACTTCTTTATCCTCTCTCCAGAGGC-TTTTTTTTCTTGCTGTTAATTT 1443
Db |||||||
458 AAATCCTAACTTCTTTATCCTCTCTCCAGAGGCTTTTTTTTCTTGCTGTTAATTT 399
QY 1444 AACATTTTAAACGCAGATATTTTGTCAAGGGGCTTGCATTCAAACCTGCTTTCCAGG 1503
Db |||||||
398 AACATTTTAAAGAGCAGATATTTTGTCAAGGGGCTTGCATTCAAACCTGCTTTCCAGG 339
QY 1504 GGTATCTCAGAGAAAGATATAAGTGTGATCTAAGAAAAAGTGATGTTTAGGAAAGT 1563
Db |||||||
338 GGTATCTCAGAGAAAGATATAAGTGTGATCTAAGAAAAAGTGATGTTTAGGAAAGT 279
QY 1564 GAAAAATATTTTGTGTTTGTATTTGAAGAAAGATGATGCTATTTTGACAGAAATCATATA 1623
Db |||||||
278 GAAAAATATTTTGTGTTTGTATTTGAAGAAAGATGATGCTATTTTGACAGAAATCATATA 219
QY 1624 TGTATGGATATATTTTAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATATAAAT 1683
Db |||||||
218 TGTATGGATATATTTTAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATATAAAT 159
QY 1684 AAAAGAGCAGAAAAATATGCTTGGTTTTCATTGCTTACCAAAAAACAAACAAAAA 1743
Db |||||||
158 AAAAGAGCAGAAAAATATGCTTGGTTTTCATTGCTTACCAAAAAACAAACAAAAA 99
QY 1744 AAGTTGTCCTTTGAGAACTTCACCTGCTCTATGTTGGTACCTGAGTCAAAATTTGTCATT 1803
Db |||||||
98 AAGTTGTCCTTTGAGAACTTCACCTGCTCTATGTTGGTACCTGAGTCAAAATTTGTCATT 39
QY 1804 TTTGTTCTGTGAAAAATAAATTTCTCTTCTTGACCAATT 1841
Db |||||||
38 TTTGTTCTGTGAAAAATAAATTTCTCTTCTTGACCAATT 1

RESULT 11
CB851224
LOCUS CB851224
DEFINITION UI-CF-DU1-adn-j-21-0-UI.s1 UI-CF-DU1 Homo sapiens cDNA clone
UI-CF-DU1-adn-j-21-0-UI 3', mRNA sequence.
ACCESSION CB851224
VERSION CB851224.1 GI:30045993
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 551)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=No.

FEATURES
source Location/Qualifiers
1..551
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-DU1-adn-j-21-0-UI"
/tissue_type="Primary Lung Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-DU1"
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-CF-DU1 is a normalized cDNA library containing the following tissue(s): Primary Lung Epithelial Cells The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GGCTGTAGGC.
TAG_SEQ=None found"

ORIGIN
Query Match 26.1%; Score 525.2; DB 14; Length 551;
Best Local Similarity 99.4%; Pred. No. 7.2e-95;
Matches 527; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 785 TTTTAACTTTACTATAAGCCATGCAAAATGACAAAAATCTATATTACTTTCTCAAAAT 844
Db |||||||
22 TTTTAACTTTACTATAAGCCATGCAAAAGACAAAATCTATATTACTTTCTCAAAAT 81
QY 845 GGACCCCAAGAAACTTTTGATTTACTGTTCTTAAGTGCCTAATCTTAATTACAGGAATG 904
Db |||||||
82 GGACCCCAAGAAACTTTTGATTTACTGTTCTTAAGTGCCTAATCTTAATTACAGGAATG 141
QY 905 TGCATCAGCTATTATGATTTCTATAAGCTATTTCAGCAGAAATGAGATATTAAACCCCAATG 964
Db |||||||
142 TGCATCAGCTATTATGATTTCTATAAGCTATTTCAGCAGAAATGAGATATTAAACCCCAATG 201
QY 965 CTTTGATTGTTCTAGAAAGTATAGTAATTTGTTTCTAAGGTGGTCAAGCATCTACTCT 1024
Db |||||||
202 CTTTGATTGTTCTAGAAAGTATAGTAATTTGTTTCTAAGGTGGTCAAGCATCTACTCT 261
QY 1025 TTTTATCATTTACTTCAAAATGACATTGCTAAAGACTGCATTATTTTACTACTGTAATTT 1084
Db |||||||
262 TTTTATCATTTACTTCAAAATGACATTGCTAAAGACTGCATTATTTTACTACTGTAATTT 321
QY 1085 CTCCACGACATAGCATTATGTACATAGATGAGTGTAAACATTTATATCTCACATAGACACA 1144
Db |||||||
322 CTCCACGACATAGCATTATGTACATAGATGAGTGTAAACATTTATATCTCACATAGACACA 381

QY 1145 TGCTTATATGGTTTTATTTAAATGAAATGCCAGTCCATTACACTGAATAAATAGAACTC 1204
Db |||||||
382 TGCTTATATGGTTTTATTTAAATGAAATGCCAGTCCATTACACTGAATAAATAGAACTC 441
QY 1205 AACTATTGCTTTTCAGGGAATCATGGATAGGTTGAAGAAGGTTACTATTATTTGTTTA 1264
Db |||||||
442 AGCTATTGCTTTTCAGGGAATCATGGATAGGTTGAAGAAGGTTACTATTATTTGTTTA 501
QY 1265 AAAACAGCTTAGGGATTAAATGCTCTCCATTTTATATGAAGATTAAAAATGA 1314
Db |||||||
502 AAAATAGCTTAGGGATTAAATGCTCTCCATTTTATATGAAGATTAAAAATGA 551

RESULT 12
BG957666/c
LOCUS BG957666 586 bp mRNA linear EST 12-JUN-2001
DEFINITION PM0-CT0802-050301-002-c03 CT0802 Homo sapiens cDNA, mRNA sequence.
ACCESSION BG957666
VERSION BG957666.1 GI:14375837
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 586)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM0&t2=PM0-CT0802-050301-002-c03&t3=2001-03-05&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 586.
Location/Qualifiers
1..586
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="CT0802"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

FEATURES
source
1..586
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="CT0802"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN
Query Match 25.6%; Score 514.4; DB 12; Length 586;
Best Local Similarity 98.5%; Pred. No. 1e-92;
Matches 540; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

QY 650 TTTTGTGCAACGAAAGAGCAGTAGCTACAGATATCTC-GATACCTTCCCATCGCACA 708
|||||

Db 576 TTTTGGTTGCAACGAAAGCGCAGTAGCTACAGATACCTCCCTTCGCACA 517

Qy 709 ACCCAAAAAAGTTATCACACCGGAAGAGTCCACGAGGTCTACTCCAGAAGTCAGTAT 768

Db 516 ACCCAAAAAAGTTATCACACCGGAAGAGTCCACGAGGTCTACTCCAGAAGTCAGTAT 457

Qy 769 GTGTAGTTGTGTATG-TTTTTTAACTTTACTATAAAGCATGCAAAATGACAAAAATCTA 827

Db 456 GTGTAGTTGTGTATGTTTTTTAACTTTACTATAAAGCATGCAAAATGACAAAAATCTA 397

Qy 828 TATTACTTTCTCAAAATGGACCCCAAGAAACTTTTGATTTACTGTTCTTAAGTGCCTAAT 887

Db 396 TATTACTTTCTCAAAATGGACCCCAAGAAACTTTTGATTTACTGTTCTTAAGTGCCTAAT 337

Qy 888 CTTAATTACAGGAAGTGTGCATCAGCTATTTATGATTTCTATAAGCTATTTTCAGCAGAATG 947

Db 336 CTTAATTACAGGAAGTGTGCATCAGCTATTTATGATTTCTATAAGCTATTTTCAGCAGAATG 277

Qy 948 AGATATTAAACCCCAATGCTTTGATTTCTAGAAAGTATAGTAATTTGTTTCTAAGGTG 1007

Db 276 AGATATTAAATCCAATGCTTTGATTTCTAGAAAGTATAGTAATTTGTTTCTAAGGTG 217

Qy 1008 GTTCAAGCATCTACTCTTTTATCAATTTACTTCAAAATGACATTGCTTAAGACTGCATTA 1067

Db 216 GTTCAAGCATCTACTCTTTTATCAATTTACTTCAAAATGACATTGCTTAAGACTGCATTA 157

Qy 1068 TTTTACTACTGTAATTTCTCCACGACATAGCATTATGTACATAGATGAGTGTAACATTTA 1127

Db 156 TTTTACTACTGTAATTTCTCCACGACATAGCATTATGTACATAGATGAGTGTAACATTTA 97

Qy 1128 TATCTCACATAGACATGCTTTATATGTTTATTTAAATGAAATGCCAGTCCATTACA 1187

Db 96 TATCTCACATAGACATGCTTTATATGTTTATTTAAATGAAATGCCAGTCCATTACA 37

Qy 1188 CTGAATAA 1195

Db 36 CTGATAAA 29

RESULT 13

AW237740/c

LOCUS xm81a03.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2690572 3',

DEFINITION AW237740 513 bp mRNA linear EST 13-DEC-1999

ACCESSION mRNA sequence.

VERSION AW237740

KEYWORDS AW237740.1 GI:6570129

SOURCE EST.

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 513)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco

High quality sequence stop: 462.

Location/Qualifiers

1. .513

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

FEATURES

source

/clone="IMAGE:2690572"

/lab_host="DH10B"

/clone_lib="NCI_CGAP_Kid11"

/note=Torgan: kidney; Vector: pUT3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 25.5%; Score 512; DB 10; Length 513;

Best Local Similarity 100.0%; Pred.No. 3.2e-92;

Matches 512; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1499 CCAGGGCTATCTCAGAAAGATAAAAGTGTGATCTAAGAAAAAGTGTGTTTAGG 1558

Db 513 CCAGGGCTATCTCAGAAAGATAAAAGTGTGATCTAAGAAAAAGTGTGTTTAGG 454

Qy 1559 AAAGTGAATAATTTTGTGTTTGAAGAGATGATGCAATTTTGACAAGAAATC 1618

Db 453 AAAGTGAATAATTTTGTGTTTGAAGAGATGATGCAATTTTGACAAGAAATC 394

Qy 1619 ATATATGATGATATATTTTAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATA 1678

Db 393 ATATATGATGATATATTTTAATAAGTATTTGAGTACAGACTTTGAGGTTTCATCAATA 334

Qy 1679 TAAATAAAGAGCAGAAAAATATGCTGTTGTTTCAATTTGCTTACCAAAAAACAAC 1738

Db 333 TAAATAAAGAGCAGAAAAATATGCTGTTGTTTCAATTTGCTTACCAAAAAACAAC 274

Qy 1739 AAAAAAGTTGCTCTTTGAGAACTTCACCTGCTCCTATGTGGGTACCTGAGTCAAAATG 1798

Db 273 AAAAAAGTTGCTCTTTGAGAACTTCACCTGCTCCTATGTGGGTACCTGAGTCAAAATG 214

Qy 1799 TCATTTTGTCTGTGAAAAATAAATTCCTTCTTGACCATTCTCTGTTAGTTTACTA 1858

Db 213 TCATTTTGTCTGTGAAAAATAAATTCCTTCTTGACCATTCTCTGTTAGTTTACTA 154

Qy 1859 AAATCTGTAATACTGTATTTTCTGTTTATTCCTTATTCCTGAACTGAACTGCAATCCAAT 1918

Db 153 AAATCTGTAATACTGTATTTTCTGTTTATTCCTTATTCCTGAACTGAACTGCAATCCAAT 94

Qy 1919 TTGAAAGTTTGTGTCGACGCTGCTGCTAGCTTAAATGATGTGTTCTATTGCTTTTATACA 1978

Db 93 TTGAAAGTTTGTGTCGACGCTGCTGCTAGCTTAAATGATGTGTTCTATTGCTTTTATACA 34

Qy 1979 TTTATATTAAATAAATTGTACATTTTCTAATT 2010

Db 33 TTTATATTAAATAAATTGTACATTTTCTAATT 2

RESULT 14

BF675808

LOCUS BF675808

DEFINITION 602083725F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4247718 5',

ACCESSION mRNA sequence.

VERSION BF675808.1 GI:11949703

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 689)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC1065 row: b column: 07
High quality sequence stop: 676.
Location/Qualifiers

FEATURES
source
1. .689
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4247718"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH MGC 83"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgctcgcc); Site 2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGGCGCGAGCGCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

ORIGIN
Query Match 25.1%; Score 505.2; DB 10; Length 689;
Best Local Similarity 95.8%; Pred. No. 6.9e-91;
Matches 638; Conservative 0; Mismatches 13; Indels 15; Gaps 11;
QY 943 GAATGAGATATTAAACCAATGCTTTTGATTTCTAGAAAGTATAGTAATTTGTTTCTA 1002
Db 1 GAATGAGATATTAAACCAATGCTT--GATGTTCTAGAAAGTATAGTAA--TTGTTTCTA 57
QY 1003 AGGTGTTCAAGCATCTACTCTTTTATCAATTACTTCAAATGACATTGCTAAAGACTG 1062
Db 58 AGGTGTTCAAGCATCTACTCTTTTATCAATTACTTCAAATGACA-TGCTAAAGACTG 116
QY 1063 CATTATTTTACTACTGTAAATTTCTCCACGACATAGCATTATGTACATAGATGAGTTAAC 1122
Db 117 CATTATTTTACTACTGTAAATTTCTCCACGACATAGCATTATGTACATAGATGAGTTAAC 176
QY 1123 ATTTATATCTCACATAGACATGCTTTATATGGTTTATTTTAAATGAAATGCCAGTCCA 1182
Db 177 ATTTATATCTCACATAGACATGCTTTATATGGTTTATTTTAAATGAAATGCCAGTCCA 236
QY 1183 TTACACTGAATAAATAGAACTCAACTATTGCTTTTCAGGAAATCATGGATGGTTGAA 1242
Db 237 TTACACTGAATAAATAGAACTCAACTATTGCTTTTCAGGAAATCATGGATGGTTGAA 296
QY 1243 GAAGGTACTATTAAATGTTTAAACACAGCTTAGGGATTAAATGCTCTCCATTATAATGA 1302
Db 297 GAAGGTACTATTAAATGTTTAAACACAGCTTAGGGATTAAATGCTCTCCATTATAATGA 356
QY 1303 AGATTAAATGAAGGCTTTAATCAGCATTTGTAAGGAAATTTGAATGGCTTTCTGATATGC 1362
Db 357 AGATTAAATGAAGGCTTTAATCAGCA-TGTAAGGAAATTTGAATGG-TTCTGATATGC 414
QY 1363 TGTTTTTTAGCCTAGGAGTTAGAAATCCTAATCTTCTTATCCTCTTCTCCAGAGGCTTT 1422
Db 415 TGTTTTTTAGCCTAGGAGTTAGAAATCCTAA-TTCTTTATCCTCTTCTCCAGAGGCTTT 473
QY 1423 TTTTCTTCT-TGTGTATTAAATTAACATTTT-AAAACGCAGATATTTTGTCAAGGGGCTT 1480
Db 474 TTTTCTTCTGTGTATTAAATTAACATTTTAAAGAGCAGATATTTGGTCAAGGGGCTT 533
QY 1481 TGCATTCAAACACTGCTTTTCCAGGGCTATACCTCAGAGAAAGATAA-AAGTGTGATCTAAG 1539

Db 534 GTGATTCAAACACTGCTTTTCCAGGGCTATACTCAGAAGAAAGATAACACCTGTGATCTAAG 593
QY 1540 AAAAAGTGATGGTTTATAGAAAGTGAATAATATTTTGTGTTTATTTGAAGAAGAATGA 1599
Db 594 AAAAAGTGATGG-TTTAGGAAAGTGAATAATATT-----TGTTTGTATTGACAGAAGAATGA 648
QY 1600 TGCATT 1605
Db 649 TGCTTT 654
RESULT 15
AY400667
LOCUS
DEFINITION
Mus musculus CLDN8 gene, VIRTUAL TRANSCRIPT, partial sequence,
Genomic survey sequence.
ACCESSION
AY400667
VERSION
AY400667.1 GI:39756656
KEYWORDS
GSS.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 678)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL
Science 302 (5652), 1960-1963 (2003)
PUBMED
14671302
REFERENCE
2 (bases 1 to 678)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE
Direct Submission
JOURNAL
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Location/Qualifiers
source
1. .678
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>678
/gene="CLDN8"
/locus_tag="HCM0633"
ORIGIN
Query Match 24.6%; Score 494; DB 29; Length 678;
Best Local Similarity 83.0%; Pred. No. 1.2e-88;
Matches 563; Conservative 0; Mismatches 115; Indels 0; Gaps 0;
QY 97 ATGGCAACCCATGCCTTAGAAATCGCTGGGCTGTTTCTTTGGTGGTGGATGGTGGC 156
Db 1 ATGGCAACCTACGCTCTTCAAATGGCTGCACTGGTGGTGGTGGTGGTGGTGGC 60
QY 157 ACAGTGGCTGTCACTGTTCATGCCTCAGTGGAGAGTGTGGCCCTTCATTGAAAACAACATC 216
Db 61 ACGTGGCTGTCACTATCATATGCCTCAGTGGAGAGTGTGCCTTCATCGAAGTAACATT 120
QY 217 GTGGTTTTTGAAAACTTCTGGGAAGGACTGTGGATCAATTGCGTGAGGAGGCTAACATC 276
Db 121 GTGGTGTTTGAAACCGCTGGGAAGGCTTGTGGATGAATTGTATGAGGCATGCCAACATC 180
QY 277 AGGATGCAGTGCAAAATCTATGATCCCTGCTGGCTCTTCTCCGGACCTACAGGCAGCC 336
Db 181 AGAATGCAGTGCAAGGTCTACGACTCCCTGCTGGCTCTTAGTCCAGACCTCCAGGCATCC 240
QY 337 AGAGGACTGATGTGTGCTGCTTCCGTGATGTCTTCTTGGCTTTTCATGATGCCCATCCTT 396

Db	241	CGAGGACTGATGTGCTGCGTCCGTTCTTGGCTTTCTAATGACAGCCATCCTC	300
Qy	397	GGCATGAAATGACACAGGTGCACGGGGACAATGAGAGGTGAAGGCTCACATTTCTGCTG	456
Db	301	GGAATGAAGTGACACAGATGCACGGGGACGATGAGAACGTGAAGAGCGGCATCTTGCTG	360
Qy	457	ACGGCTGGAATCATCTTTCATCATCACGGGCATGGTGGTCTCATCCCTGTGAGCTGGGTT	516
Db	361	ACAGCCGGAATCATCTTCTTCATCACCGGCTTGGTTGTGCTCATCCCTGTGAGCTGGGTT	420
Qy	517	GCCAAATGCCATCATCAGAGATTTCTATACTCAATAGTGAATGTTGCCAAAACGTGAG	576
Db	421	GCCAAATCCATCATCAGAGACTTCTACAAACCCACTGGTGGATGTGGCCCTAAAGCGCGAG	480
Qy	577	CTTGAGAGAGCTCTCTACTTAGGATGGACCAACGGCACTGGTCTGATGTTGGAGGAGCT	636
Db	481	CTGGGAGAGGCCCTCTACATAGGCTGGACCAACAGCGCTGGTCTGATCGCTGGAGGAGCA	540
Qy	637	CTGTTCTGCTGCGTTTTTTTGTGCAACGAAAAGAGCAGTAGCTACAGATACTCGATACCT	696
Db	541	CTGTTCTGTTGTGTTTTTTGTTGTAAGGAGGAGCAACAGTTACAGGTAAGTCTCGGTACCA	600
Qy	697	TCCCATCGCACAAACCCAAAAGTTATCACACCGGAAAGAGTCAACCGAGCGTCTACTCC	756
Db	601	TCCCATCGCACCACTCAACGGAGTTTCCACGCCGAAAAGAGATCTCCGAGCATATACTCC	660
Qy	757	ACAAGTCAGTATGTGTAG 774	
Db	661	AAAAGTCAGTATGTGTAG 678	

RESULT 16
AW205135/c
LOCUS
DEFINITION UI-H-BI1-aem-g-01-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens CDNA clone
IMAGE:2720040 3', mRNA sequence.
AW205135
VERSION AW205135.1 GI:6504607
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 492)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.
FEATURES
source
1. .492
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2720040"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP Sub3"
/notes="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The
NCI_CGAP_Sub3 library is a subtracted library derived from
the NCI_CGAP_Sub1 library, which is a subtracted library
derived from BI. BI constitutes a mixture of 21

normalized or subtracted NCI_CGAP libraries:
NCI_CGAP_Co4, NCI_CGAP_Pr22, NCI_CGAP_Pr28, NCI_CGAP_Co10,
NCI_CGAP_Co16, NCI_CGAP_Kid5, NCI_CGAP_Kid12,
NCI_CGAP_Kid3, NCI_CGAP_Kid11, NCI_CGAP_Lym2,
NCI_CGAP_Br2, NCI_CGAP_Co8, NCI_CGAP_CLL1, NCI_CGAP_Lei2,
NCI_CGAP_Brn23, NCI_CGAP_Lu5, NCI_CGAP_Lu24,
NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6,
NCI_CGAP_Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI_CGAP_Kid3 pool 1 LLM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clones 1322376-1323911,
1456008-1456775, 1500552-1502855); NCI_CGAP_Kid5 pool 1
LLM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clones
1323912-1325831, 1471368-1472903, 1492104-1493255);
NCI_CGAP_Lu5 pool 1 LLM 3575-3582, 3851-3854 (IMAGE
Clones 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
pool 1 LLM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clones 1257096-1258631, 1469064-1470983,
1475592-1476743); NCI_CGAP_Pr22 pool 1 LLM 2457-2459,
2758-2759, 3062-3068 (IMAGE Clones 985608-986759,
1101192-1101959, 1217928-1220615); NCI_CGAP_Co10 pool 1
LLM 2644-2653, 2871-2872 (IMAGE Clones 1057416-1061255,
1144584-1145351). Subtraction was performed as previously
described [Bonaldo, Lennon & Soares (1996): Normalization
and Subtraction: Two Approaches To Facilitate Gene
Discovery. Genome Research 6, 791-806.
TAG_TISSUE=kidney
TAG_LIB=NCI_CGAP_Kid3
TAG_SEQ=AATGC"

ORIGIN

Qy	1219	AGGAAATCATGATAGGTTGAAGAAGGTTACTATTAATGTTTAAAAACAGCTTAGGG	1278
Db	492	AGGAAATCATGATAGGTTGAAGAAGGTTACTATTAATGTTTAAAAACAGCTTAGGG	433
Qy	1279	ATTAATGCTCTCCATTTATAATGAAGATTAAATGAAGCTTTAATCAGCATTTAAAGG	1338
Db	432	ATTAATGCTCTCCATTTATAATGAAGATTAAATGAAGCTTTAATCAGCATTTAAAGG	373
Qy	1339	AAATTGAATGGCTTCTGATATGCTGTTTTTTAGCCTAGGAGTTAGAAATCCTAATCT	1398
Db	372	AAATTGAATGGCTTCTGATATGCTGTTTTTTAGCCTAGGAGTTAGAAATCCTAATCT	313
Qy	1399	TTATCCTCTTCTCCAGAGGCTTTTTTTTCTTGTTGTTAATAATTAACATTTTAAACG	1458
Db	312	TTATCCTCTTCTCCAGAGGCTTTTTTTTCTTGTTGTTAATAATTAACATTTTAAAG	253
Qy	1459	CAGATATTTTGCAAGGGGCTTTGCAATCAAACCTGCTTTCCAGGGCTATCTCAGAAGA	1518
Db	252	CAGATATTTTGCAAGGGGCTTTGCAATCAAACCTGCTTTCCAGGGCTATCTCAGAAGA	193
Qy	1519	AAGATAAAGTGTGATCTAAGAAAAGTGATGGTTTTTAGGAAAAGTGAAAATATTTTGT	1578
Db	192	AAGATAAAGTGTGATCTAAGAAAAGTGATGGTTTTTAGGAAAAGTGAAAATATTTTGT	133
Qy	1579	TTTGTATTTGAAGAAGATGATGCAATTTTGACAAGAAATCATATATGTATCGATATATT	1638
Db	132	TTTGTATTTGAAGAAGATGATGCAATTTTGACAAGAAATCATATATGTATCGATATATT	73
Qy	1639	TAATAAGTATTTGAGTACAGACTTTGAGGTTTTCATCAATATAATAAAGACAGAAAAA	1698
Db	72	TAATAAGTATTTGAGTACAGACTTTGAGGTTTTCATCAATATAATAAAGACAGAAAAA	13
Qy	1699	TA 1700	
Db	12	TA 11	

RESULT 17
CB418866
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CB418866
591692 MARC 6BOV Bos taurus cDNA 5', mRNA linear EST 25-MAR-2003
CB418866
CB418866.1 GI:291833348
EST.
Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 683)
Smith,T.P.L., Roberts,A.J., Ehternkamp,S.E., Chitko-McKown,C.G.,
Wray,J.E. and Keele,J.W.
A second set of bovine ESTs from pooled-tissue normalized libraries
Unpublished (2003)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: FQY8011 row: A column: 23
Seq primer: GTAATACGACTCACTATAGG.
Location/Qualifiers
1. .683
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 6BOV"
/note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including liver, lung, hypothalamus, pituitary, and
placenta/endometrium."

FEATURES
source

1. .683
/organism="Bos taurus"
/mol_type="mRNA"
/strain="Hereford Shorthorn"
/db_xref="taxon:9913"
/clone="CCL003314"
/sex="female"
/tissue_type="pooled"
/dev_stage="Adult"
/lab_host="XL1-BlueMRF'strain"
/clone_lib="Bos taurus skin cDNA library"
/note="Organ: skin; Vector: Uni-ZAPXR; Site_1: EcoRI;
Site 2: Xho I; Library made from pooled skin of adult
female Hereford-Shorthorn."

ORIGIN

Query Match 23.3%; Score 468.4; DB 14; Length 683;
Best Local Similarity 83.1%; Pred. No. 1.6e-83;
Matches 570; Conservative 0; Mismatches 111; Indels 5; Gaps 3;

QY 294 CTATGATTCCTGCTGGCTCTTTCTCCGACCTACAGGCAGCCAGGAGACTGATGTGC 353
|||
Db 1 CTACGATCGCTGCTGGCTCTCTCTCCGACCTACAGGCAGCCAGGAGACTGATGTGC 60
|||
QY 354 TGCTTCGGTGATGTCCTTCTTGGCTTTTCATGATGGCCATCTTGGCATGAATGCACCAG 413
|||
Db 61 CGCCTCGGTGCTGGCTTCTTGGCTTTCCTGACGGCCGCTCCTCGGCATGAAGTGACCAG 120
|||
QY 414 GTGCACGGGGACAATGAGAAGGTGAAGGCTCACATTCTGCTGACGGCTGGAATCATCTT 473
|||
Db 121 ATGCGCGGGGACGACGACAAGGTGAAGGTCAATTCTGCTGACCGCTGGAGTGATTT 180
|||
QY 474 CATCATCAGGGGCATGGTGTGCTCATCCCTGAGCTGGGTTGCCAATGCCATCATCAG 533
|||
Db 181 CATCATCACTGGCCCTCGTGTGCTCATCCCGTGAGCTGGGTTGCCAATTCATCATCAG 240
|||
QY 534 AGATTTCTATAACTCAATAGTGAATGTTGCCAAAACGTCGAGCTGGAGAAGCTCTCTA 593
|||
Db 241 AGACTTCTACAACCCCAATAGTGATATTGCCAGAAACGGAGCTGGGAGAACCCCTCTA 300
|||
QY 594 CTTAGGATGGACACGGCACTGGTGTGCTGATTGTGGAGGAGCTCTGTTCTGTCGGTTTT 653
|||
Db 301 CATAGGCTGGACACGGCCCTGGTGTGCTGATTGTGGAGGGCGCTGTTCTGTTGCGTTT 360
|||
QY 654 TTGTTGCAACGAAAAGAGCAGTACAGATCTCGATACCTTCCCATCGCACACCCA 713
|||

Db 361 CTGTTGCCATGAAAAGAGCAGTAGTACAGATACTCCATACCGTCCCACGAAACAACCCA 420
QY 714 AAAAAGTTATCACACCGGAAAGAAAGTACCGAGCGTCTACTCCAGAAGTCAGTATGTGA 773
|||
Db 421 GAAAAGCTATCACGCGCGAAAGAAAGTCCGAGTGTGTACTCCAAAAGTCAGTACGTGA 480
|||
QY 774 GTTGTGTATGTTTTTTTAACTTTTACTATAAAGCCCATGCAAAATGACAAAATCTATATTAC 833
|||
Db 481 GTTATG--ACTCTTTTAACTACCTAGAAAGCCACCCAGATGACATAGATGTCCACT-C 537
|||
QY 834 TTTCTCAAAATGGACCCCAAGAAACTTTTGATTACTTGTCTTTAACTGCCTAATCTTAAT 893
|||
Db 538 TTCCCAAAACTGAGACCCCAAGAACATTGATTGGTGTCTTTAACTGCCTGGTATTAAT 597
|||
QY 894 TACAGGAACGTG--TGCATCAGCTATTTATGATTCTCTATAAGCTATTTTCAGCAGAATGAGAT 951
|||
Db 598 CACAGGACTGATTACATCAGCTCTTTATGATTCTATAATCAATTTTCAGCTGCATGAGAA 657
|||
QY 952 ATTAACCCCAATGCTTTGATTGTCT 977
|||
Db 658 ATTGTACACATTGCTTTGATTATTCT 683
|||

RESULT 18
CF766535
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CF766535
CES003314 Bos taurus skin cDNA library Bos taurus linear EST 17-OCT-2003
CCL003314 5', mRNA sequence.
CF766535
CF766535.1 GI:37715754
EST.
Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 681)
Wang,Y.H., McWilliam,S. and Lehnert,S.
Transcription profiling of cattle skin
Unpublished (2003)
Contact: Dr Yonghong Wang
Functional Genomics Lab
CSIRO Livestock Industries
Level 5, Queensland Biosciences Precinct, University of Queensland,
306 Carmody Road St.Lucia QLD Australia
Tel: 07 3214 2445
Fax: 07 3214 2685
Email: Yonghong.Wang@csiro.au
Plate: 43 row: B column: 11.
Location/Qualifiers
1. .681
/organism="Bos taurus"
/mol_type="mRNA"
/strain="Hereford Shorthorn"
/db_xref="taxon:9913"
/clone="CCL003314"
/sex="female"
/tissue_type="pooled"
/dev_stage="Adult"
/lab_host="XL1-BlueMRF'strain"
/clone_lib="Bos taurus skin cDNA library"
/note="Organ: skin; Vector: Uni-ZAPXR; Site_1: EcoRI;
Site 2: Xho I; Library made from pooled skin of adult
female Hereford-Shorthorn."

ORIGIN

Query Match 23.0%; Score 462.2; DB 14; Length 681;
Best Local Similarity 80.6%; Pred. No. 2.8e-82;
Matches 539; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 1 GGAAGAACTGTTCTCTTCTGTGGCAGAGAACCTTGCTTCAAAGCAGAGTAGCAGTTTC 60
|||
Db 13 GGAAGAACTGCTCTCTCCCTCGGGCGCAGAGAACCTTGCTTCAGAGCAGCCTTACCAGCTC 72
|||

Qy	61	CGGAGTCCAGCTGGCTAAAACTCATCCAGAGGATAATGGCAACCCATGCCCTTAGAAATC	120
Db	73	AGGCGCCTGGTCGGATAAAACCCACCTGGAGGAGATGGCTACCTACGCCCTGCAATC	132
Qy	121	GCTGGGCTGTTTCTTGGTGGTGTGGAATGGTGGGCACAGTGGCTGTCACTGTATGCCT	180
Db	133	GCCGACTGGTGGTGGTGGTGGCATGGAGGGACAGTGGCTGCCACGGTCAATGCCT	192
Qy	181	CAGTGGAGAGTGTGGCCTTCATTGAAAAACAACATCGTGGTTTTTTGAAAACTTCTGGGAA	240
Db	193	CAGTGGAGAGTGTGGCCTTCATTGAAAAACAATATGGGGTCTTTGAAAAACCTCTGGGAA	252
Qy	241	GGACTGTGGATGAATGCGTGAGGAGGCTAACATCAGGATGCAGTGCAAAATCTATGAT	300
Db	253	GGACTATGGATGAGTTGCATGAGGCATGCTTAACATCAAAATGCAGTGCAAAATCTACGAC	312
Qy	301	TCCCTGTGGCTTTTCTCCGACCTACAGGCAGCCAGGAGTGAATGTGTGTGCTTCC	360
Db	313	TCGCTGTGGCTCTCTCTCCGACCTACAGGCAGGCAGGAGTGAATGTGTGCTGCGCTCG	372
Qy	361	GTGATGTCCCTTTCTGGCTTTTCATGATGGCCATCCTTGGCATGAAATGCACCGTGCACG	420
Db	373	GTGCTGGCCTGTTTGGCTTTTCTGACGGCCGNACTCGGCATGAAGTGTACCGATGCGCC	432
Qy	421	GGGGACAATGAGAAGGTGAAGGCTCACATTTCTGTGACGGCTGGAATCATCTTTCATC	480
Db	433	GGAGACGACGACAAGGTGAAGGTCACATTTCTGTGACCGCTGGAGTATATCATCATC	492
Qy	481	ACGGCATGGTGTGCTCATCCCTGTGAGCTGGTGGTGGCAATGCCATCATCAGAGATTTC	540
Db	493	ACTGGCCTCGCGTGTCTCATCCCGAGAGCTGGTGGTGGCAATTCATCATCAGAGACTTC	552
Qy	541	TATAACTCAATAGTGAATGTTGCCAAAAAACGTGAGCTTGGAGAGCTCTCTACTTAGGA	600
Db	553	TACAACCCCAATAGTGGATATTGCCCGAGAAACGGGAGCTGTGAGAAGCCCTTACATATGC	612
Qy	601	TGGACCACGGCACTGGTCTGATTTGAGGAGCTCTGTTCTGCTGCGTTTTTTTGTTC	660
Db	613	TAGACCACTGACCTGGTCTTGCAATGTGGAGNGCGTGTCTGTTGCGTTTCTGTTGC	672
Qy	661	AACGAAAAG 669	
Db	673	CATGAAAAG 681	

RESULT 19
CF169716
LOCUS
DEFINITION
musculus cDNA clone NIA:B0817B10 IMAGE:30469077 5', mRNA sequence.
CF169716
CF169716.1 GI:33279265
EST.
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 633)
Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method
Genome Res. 11 (9), 1553-1558 (2001)
MEDLINE
21429098
PUBMED
11544199
COMMENT
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: B0817 row: B column: 10
Seq primer: M13 Reverse

High quality sequence stop: 633
POLYA=No.
FEATURES
Location/Qualifiers
1..633
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaEST:B0817B10-5"
/db_xref="taxon:10090"
/clone="NIA:B0817B10 IMAGE:30469077"
/dev_stage="Newborn Kidney"
/lab_host="DH10B"
/clone_lib="NIA Mouse Newborn Kidney cDNA Library (Long
1)"
/note="Vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI;
Site_2: NotI; Mouse cDNA project by the Laboratory of
Genetics, National Institute on Aging (NIA), Intramural
Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA).
In brief, double-stranded cDNAs were synthesized with an
Oligo(dT) primer [Invitrogen:
5'-pGACTAGTTCTAGTCGAGCGGCCCTTTTTTTTTTTT-3'] from
26 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to Lona-linker LL-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pCMV-SPORT6 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by the standard chemical method. The
average insert size is about 3.0 kb. The library was
constructed by Yulian Piao."

ORIGIN

Query Match	22.9%;	Score 460.2;	DB 14;	Length 633;
Best Local Similarity	82.9%;	Pred. No. 7.2e-82;		
Matches	525;	Conservative	0;	Mismatches 108; Indels 0; Gaps 0;
Qy	98	TGGCAACCCATGCCTTAGAAATCGCTGGGCTGTTTCTTGGTGGTGGTGGGCA	157	
Db	1	TGGCAACCTACGCTCTTCAAATGGCTGCACCTGGTGGTGGTGGGCA	60	
Qy	158	CAGTGGCTGTCACTGTCAATGCCTCAGTGGAGAGTGTGGCCCTTCATTGAAACACATCG	217	
Db	61	CGGTGGCTGTGACTATCATGCCTCAGTGGAGAGTGTGCTTTCATCGAAAGTAACATTG	120	
Qy	218	TGGTTTTTGAACCTTCTGGGAAGGACTGTGGATGAATTCGCTGAGGCAGGCTAACATCA	277	
Db	121	TGGTGTGTTGAGAACCGCTGGGAAGGCTTGTGGATGAATTCATGAGGCATGCCAACATCA	180	
Qy	278	GGATGCAGTGCAAAATCTATGATTCCTGCTGGCTCTTCTCCGGACCTCAGGCAGCCA	337	
Db	181	GAATGCAGTGCAAGTCTACGACTCCCTGCTGGCTCTTAGTCCAGACCTCCAGGCATCCC	240	
Qy	338	GAGGACTGATGTGCTGCTTCCGTCGATGTCCTTCTTGGCTTTCATGATGGCATCCTTG	397	
Db	241	GAGGACTGATGTGCTGCTGCTCCGTCCTGCTTCTTGGCTTTCATGATGGCATCCTTG	300	
Qy	398	GCATGAAATGCACAGGTGCACGGGGACCAATGAGAAAGTGAAGGCTCACATCTGCTGA	457	
Db	301	GAATGAAGTGACCAAGATGCACGGGGACCGATGAGAACGTGAAGAGCCGCTTGTGCTGA	360	
Qy	458	CGGCTGGAATCATCTTTCATCATCAGCGGCATGGTGGTGGTGGTGGTGGTGGTGGTGG	517	
Db	361	CAGCCGGAATCATCTTCTTCATCACCCTGCTGGTGTGCTCATCCCTGTGAGTGGGTTG	420	
Qy	518	CCAATGCCATCATCAGAGATTTCTATAACTCAATAGTGAATGTTGCCCAAAACGTGAGC	577	
Db	421	CCAATGCCATCATCAGAGATTTCTATAACTCAATAGTGAATGTTGCCCAAAACGTGAGC	480	
Qy	578	TTGGAAGACTCTCTACTTAGGATGGACCGCACTGGTGTGATTTGTTGGAGGAGCTC	637	

Db 481 TGGGAGAGCCCTCTACATAGGCTGGACCACAGCGCTGGTGTGATCGCTGGAGGACAC 540
QY 638 TGTTCTGCTGCGTTTTTTTGTGCAACGAAAGAGCAGTAGCTACAGATACTCGATACCTT 697
Db 541 TGTTCTGTTGTGTTTGTGTACTGAAAGGAGCAACAGTTACAGGTACTCGGTACCAT 600
QY 698 CCCATCGCACAAACCCAAAAGTTATCACACCG 730
Db 601 CCCATCGCACCACTCAACGGAGTTTCCACGCCG 633

RESULT 20
BG402412 809 bp mRNA linear EST 12-MAR-2001
LOCUS 602466111F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4594155 5',
DEFINITION mRNA sequence.
ACCESSION BG402412
VERSION BG402412.1 GI:13295860
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 809)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM1335 row: e column: 04
High quality sequence stop: 567.
Location/Qualifiers
1. 809
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4594155"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_75"
/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgctcgcc); Site_2: SfiI (ggccattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGAGCGGCCGACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.65
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."

ORIGIN
Query Match 22.1%; Score 445.2; DB 12; Length 809;
Best Local Similarity 96.1%; Pred. No. 6.9e-79;
Matches 490; Conservative 0; Mismatches 13; Indels 7; Gaps 3;

QY 1 GGAAAACTGTTCTCTTCTGTGGCAGAGAACCCCTGCTTCAAACGAGAGTAGCAGTTC 60
Db 15 GGAAAACTGTTCTCTTCTGTGGCAGAGAACCCCTGCTTCAAACGAGAGTAGCAGTTC 74
QY 61 CGGAGTCCAGCTGGCTAAACTATCCCGAGAGATAATGGCAACCCATGCCCTTAGAAATC 120
Db 75 CGGAGTCCAGCTGGCTAAACTATCCCGAGAGATAATGGCAACCCATGCCCTTAGAAATC 134
QY 121 GCTGGGCTGTTTCTTGTGGTGTGGAAATGGTGGGCACAGTGGCTGTCACTGTATGCCT 180

Db 135 GCTGGGCTGTTTCTTGTGGTGTGGAAATGGTGGGCACAGTGGCTGTCACTGTATGCCT 194
QY 181 CAGTGGAGAGTGTGGCCCTTCATTGAAAAACAACATCGTGGTGTTTTGAACAACTTCTGGAA 240
Db 195 CAGTGGAGAGTGTGGCCCTTCATTGAAAAACAACATCGTGGTGTTTTGAACAACTTCTGGAA 254
QY 241 GGAAGTGGATGAATTGCGTGAGSCAGGCTAAACATCAGGATGCAGTGCAAAATCTATGAT 300
Db 255 GGAAGTGGATGAATTGCGTGAGSCAGGCTAAACATCAGGATGCAGTGCAAAATCTATGAT 314
QY 301 TCCCTGCTGGCTCTTTCTCGGACCTACAGGCAGCCAGAGGACTGATGTGTCTGCTTCC 360
Db 315 TCCCTGCTGGCTCTTTCTCGGACCTACAGGCAGCCAGAGGACTGATGTGTCTGCTTCC 374
QY 361 GTGATGTCCTTCTTGGCTTTCA-TGATGGCCATCCTTGG----CATGAATGCACCAGGT 415
Db 375 GTGATGTCCTTCTTGGCTTTCACTGATGGCCATCCTTGGGCTGTGAACATGCACCAGGT 434
QY 416 GCACGGGGGACAAATGAGAAAGGTGAAGGCTCACATTCCTGCTACGGCTGGAAATCA--TCTT 473
Db 435 GCACGGGGGACGATGAGAAAGGTGAAGGCTCACATTCCTGCTACGGCTGGAAATCACTCTTT 494

RESULT 21
BI852250 902 bp mRNA linear EST 10-OCT-2001
LOCUS 603378454F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5391109 5',
DEFINITION mRNA sequence.
ACCESSION BI852250
VERSION BI852250.1 GI:15992997
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 902)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM1996 row: k column: 14
High quality sequence stop: 826.
Location/Qualifiers
1. 902
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:5391109"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam2"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

ORIGIN
Query Match 22.0%; Score 443.2; DB 12; Length 902;
Best Local Similarity 80.4%; Pred. No. 1.7e-78;

Matches 569; Conservative 0; Mismatches 133; Indels 6; Gaps 4;

Qy 90 GAGGATAATGGCAACCCCATGCCCTTATAGAAATCGCTGGCTGTTCTTGTGGTGTGGAAAT 149
|||||
Db 137 GAGGACGATGGCAACCTACGCTCTTCAAAATGGCTGCACTGGTGTGGTGTGGCAT 196
|||||

Qy 150 GGTGGGCACAGTGGCTGTCACTGTCTCATGCTCCTCAGTGGAGAGTGTGCGCTTCATTGAAAA 209
|||||
Db 197 GGTGGGCACGGTGGCTGTGACTATCATGCCTCAGTGGAGAGTGTCTGCCTTCATCGAAAG 256
|||||

Qy 210 CAACATCGTGGTTTTTGAAGAACTTCTGGGAAGGACTGTGGATGAATTGCGTGAGGCAGGC 269
|||||
Db 257 TAACATTGTGGTTTGAGAACCGCTGGGAAGGCTTGTGGATGAATTGTATGAGGCATGC 316
|||||

Qy 270. TAACATCAGGATGAGTGCAAAATCTATGATTCCCTGCTGGCTCTTCTCCGGACCTACA 329
|||||
Db 317 CAACATCAGAAATGAGTGCAAGGTCTACGACTCCCTGCTGGCTCTTAGTCCAGACCTCCA 376
|||||

Qy 330 GGCAGCCAGAGGACTGATGTGTGCTGCTTCCGTGATGTCCTTCTTGGCTTTTCATGATGC 389
|||||
Db 377 GGCATCCCGAGGACTGATGTGTGCTGCTGCTGCTTCTTGGCTTTCTTGGCTTTTCATGACAGC 436
|||||

Qy 390 CATCCTTGGCATGAAATGCACCAAGTGCACGGGGGACAAATGAGAAGTGAAGGCTCACAT 449
|||||
Db 437 CATCCTCGGAATGAAGTGCAACAGATGCACGGGGGACGATGAGAACGTAAGACCGCAT 496
|||||

Qy 450 TCTGCTGACGGCTGGAAATCATCTTTCATCATCACGGGGATGGTGGTCTCATCCCTGTGAG 509
|||||
Db 497 CTTGCTGACAGCCGGAATCATCTTCTTCATCACCGGCTGGTGGTGTCTCATCCCTGTGAG 556
|||||

Qy 510 CTGGGTTGCCAATGCCATCATCAGAGATTTCTATACTCAATAGTGAATGTTGCCCAAAA 569
|||||
Db 557 CTGGGTTGCCAATTCATCATCAGAGACTTCTACACCCACTGGTGGATGTGGCCCTAAA 616
|||||

Qy 570 ACGTGAGCTTGGAGAACTCTCTACTAGGATGGACACGGCACTGGTGTCTGAT - TGTTG 628
|||||
Db 617 GCGGAGCTGGGAGAAAGCCCTCTACATAGGCTGGACACAGCGCTGGTGTCTGATCCGCTG 676
|||||

Qy 629 GAGGAGCTCTGTTCTGCTGCGTTTTTTTGTGCAACGAAAAGAGCAGTAGCTACAGATACT 688
|||||
Db 677 GAGGAGCACTGTTCTGTAGTGTGACCGTTGTACTGAAAGGAGCAACAGTTACAGGTACT 736
|||||

Qy 689 CGATACC-TTCCCATCGCACAAACCCAAAAGTTATACACA---CCGGAAGAGTCAACCG 744
|||||
Db 737 CGGTACCATCCCATCGGACCACTCAACGGGAAGTTTCAACGCGCAAAAGAGATCTCCG 796
|||||

Qy 745 AGCGTCTACTCCAGAAAGTCAGTATGTAGTTGTGTATGTTTTTTTAA 792
|||||
Db 797 AGCATATTCTCCAAAAGTCAG-ATGTGAAGTGTGTATGATATAATTTA 843
|||||

RESULT 22
BF231401
LOCUS BF231401 557 bp mRNA linear EST 27-MAR-2003
DEFINITION 253787 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF231401
VERSION BF231401.1 GI:11170371
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 557)
Sonstegard,T., Capuco,A.V., White,J., Van Tassell,C.P.,
Connor,E.E., Cho,J., Sultana,R., Shade,L., Wray,J.E., Wells,K.D.
and Quackenbush,J.
Analysis of bovine mammary gland EST and functional annotation of
the Bos taurus gene index
Mamm. Genome 13 (7), 373-379 (2002)
22135956
12140684
COMMENT Contact: Sonstegard TS

USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414
Email: tads@psai.barc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.

PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCCGAGTCACGACG
Plate: 105 row: I column: 20
Seq primer: ATTAGGTGACACTATAG.

FEATURES
source
1. .557
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="BARC 5BOV"
/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled mRNA isolated from mammary
tissues at eight physiological, developmental, and disease
states."

ORIGIN
Query Match 21.9%; Score 440.2; DB 10; Length 557;
Best Local Similarity 86.9%; Pred. No. 7.5e-78;
Matches 484; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 204 TGAAGCAATATTGTGGTCTTTTGAAGACCTCTGGGAAGGACTATGGATGAGTGCATGAG 263
|||||
Db 1 TGAAGCAATATTGTGGTCTTTTGAAGACCTCTGGGAAGGACTATGGATGAGTGCATGAG 60
|||||

Qy 264 GCAGGCTAACATCAGGATGCAGTGCAAAATCTATGATCCCTGCTGGCTCTTTCTCCGGA 323
|||||
Db 61 GCATGCTAACATCAGAAATGCAGTGCAAAATCTACGACTCGCTGCTGCTCTCTCTCCGGA 120
|||||

Qy 324 CCTACAGGCAGCCAGAGGACTGATGTGTGCTGCTTCCGATGTCCTTCTGGCTTTTCAT 383
|||||
Db 121 CCTACAGGCAGCCAGAGGACTGATGTGTGCTGCTTCCGATGTCCTTCTGGCTTTTCCT 180
|||||

Qy 384 GATGGCCATCCTTGGCATGAAATGCACAGGTCACGGGGACAATGAGAAGGTGAAGGC 443
|||||
Db 181 GACGGCCGTCCTCGCATGAAGTGTACAGATGCGCCGGGGACGACGACAAGGTGAAAGG 240
|||||

Qy 444 TCACATTCTGCTGACGGCTGGAATCATCTTTCATCATCACGGGCATGGTGGTCTCATCCC 503
|||||
Db 241 TCACATTCTGCTGACGGCTGGAGTGTTCATCATCATCTGGCCTCGTGGTCTCATCCC 300
|||||

Qy 504 TGTGAGCTGGGTTGSCAATGCCATCATCAGAGATTTCTATAACTCAATAGTGAATGTTGC 563
|||||
Db 301 CGTGAGCTGGGTTGCCAATTCATCATCAGAGACTTCTACACCCCAATAGTGGATATTGC 360
|||||

Qy 564 CCAGAAACGTCAGCTGGAGAGCTCTCTACTTAGGATGGACCCACGGCAGCTGGTCTGAT 623
|||||
Db 361 CCAGAAACGTCAGCTGGAGAGCCCTCTACATAGGCTGGACCCACGGCCTGGTCTGAT 420
|||||

Qy 624 TGTGGAGGAGCTCTGTTCTGCTGGCTTTTGTGTCACGAAAAGAGCAGTAGCTACAG 683
|||||
Db 421 TGTGGAGGAGGCTGTTCTGTTGGCTTTCTGTTGCTTGTGCTGAAAGAGCAGTAGCTACAG 480
|||||

Qy 684 ATACTCGATACCTTCCCATCGCACACCCAAAAGTTATCACACCGGAAAAGATCACC 743
|||||
Db 481 ATACTCCATACCGTCCCAACGAAACCCAGAAAGTATCACGCCGAAAAGATCGCC 540
|||||

Qy 744 GACGCTCTACTCCAGAA 760
|||||
Db 541 GAGTGTGTACTCCAAA 557
|||||

CA313163
LOCUS CA313163 486 bp mRNA linear EST 04-NOV-2002
DEFINITION UI-CF-FN0-aex-d-08-0-UI.s1 UI-CF-FN0 Homo sapiens cDNA clone
UI-CF-FN0-aex-d-08-0-UI 3', mRNA sequence.
ACCESSION CA313163
VERSION CA313163.1 GI:24531261
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 486)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
The following repetitive elements were found in this cDNA
sequence: 404-466, >AT rich#Low_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.
FEATURES
source Location/Qualifiers
1..486
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-FN0-aex-d-08-0-UI"
/tissue_type="Human Lung Epithelial cells"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-FN0"
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-FN0 is a subtracted cDNA library derived from two
normalized Human lung epithelial cell libraries (EN1 and
DUI) The library was subtracted according to according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. For additional information, contact:
bento-soares@uiowa.edu
TAG_TISSUE=Lung Epithelial Cells Tissue nos 359-368
TAG_LIB=UI-CF-FN0
TAG_SEQ=GGCTGTAGGC"
ORIGIN
Query Match 21.8%; Score 438.8; DB 14; Length 486;
Best Local Similarity 99.5%; Pred. No. 1.5e-77;
Matches 440; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1569 TATTTTGTGTTTGTATTTGAAGAAGATGATGCATTTTGACAGAAATCATATATGTAT 1628
Db 2 TTTTGTGTTTGTATTTGAAGAAGATGATGCATTTTGACAGAAATCATATATGTAT 61
QY 1629 GGATATATTTTAAGTATTGTAGTACAGACTTTGAGGTTTCATCAATATAAATAAAG 1688
Db 62 GGATATATTTTAAGTATTGTAGTACAGACTTTGAGGTTTCATCAATATAAATAAAG 121
QY 1689 AGCAGAAAAATATGCTCTGGTTTTCATTTGCTTACCAAAAAACAAACAAAAAGTT 1748
Db 122 AGCAGAAAAATATGCTCTGGTTTTCATTTGCTTACCAAAAAACAAACAAAAAGTT 181

QY 1749 GTCCTTTGAGAACTTCACCTGCTCCTATGTGGGTACTGAGTCAAAATGTCATTTTGT 1808
Db 182 GTCCTTTGAGAACTTCACCTGCTCCTATGTGGGTACTGAGTCAAAATGTCATTTTGT 241
QY 1809 TCTGTGAAAAATAAATTCCTTCTTGTGTACCAATTCCTGTTAGTTTACTAAAAATCTGTAA 1868
Db 242 TCTGTGAAAAATAAATTCCTTCTTGTGTACCAATTCCTGTTAGTTTACTAAAAATCTGTAA 301
QY 1869 ATACTGTATTTTCTGTTTATTCAAAATTTGATGAACTGACAATCCAAATTTGAAAAGTTT 1928
Db 302 ATACTGTATTTTCTGTTTATTCAAAATTTGATGAACTGACAATCCAAATTTGAAAAGTTT 361
QY 1929 GTGTCGACGCTGCTAGCTTAAATGAATGTGTTCTATTTGCTTTTATACATTTATATAA 1988
Db 362 GTGTCGACGCTGCTAGCTTAAATGAATGTGTTCTATTTGCTTTTATACATTTATATAA 421
QY 1989 TAAATTGTACATTTTCTTAATT 2010
Db 422 TAAATTGTACATTTTCTTAATT 443
RESULT 24
CF796030/c
LOCUS CF796030 688 bp mRNA linear EST 21-OCT-2003
DEFINITION 892243 MARC 4PIG Sus scrofa cDNA 3', mRNA sequence.
ACCESSION CF796030
VERSION CF796030.1 GI:37800603
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 688)
AUTHORS Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,
Nonneman,D.J., Wray,J.E. and Keele,J.W.
TITLE Porcine EST collection using a normalized library constructed from
embryos representing early developmental stages
JOURNAL Unpublished (2003)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: TMW8018 row: G column: 20
Seq primer: TAGAAGGCACAGTCGAGG.
FEATURES
source Location/Qualifiers
1..688
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 4PIG"
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with combined RNA from day-10, day-13,
day-15, day-25, and day-30 whole embryos."
ORIGIN
Query Match 21.3%; Score 429; DB 14; Length 688;
Best Local Similarity 81.4%; Pred. No. 1.3e-75;
Matches 561; Conservative 0; Mismatches 115; Indels 13; Gaps 5;
QY 732 AAAGAAGTCACCGCGCTCTACTCCAGAAGTCAGTAGTGTGTATGTTT--TT 789
Db 688 AAAGAAGTCCCGAGTGTGTACTCCAGAAGTCAGTAGTGTGTGCGTCTCTCTT 629
QY 790 TAACTTTACTATAAGCCCATGCAAAATGACAAAATCTATATTACTTTCTCAAATGGACC 849

Db	628	TGAACTGCCTAGAAAGCCACGGCTGATGAGAAAAAGGTCTACTATTTTCTAAAAATGGAAC	569
Qy	850	CCAAAGAAACTTTGATTTACTGTTCTTAACTGCCTAACTCTTAATTACAGGAACGTGTGAT	909
Db	568	CCAAAGAAACACTGATTTGCTGTTCTTGACTGCCTGACATTAATTACAGGAACCTCTTCAT	509
Qy	910	CAGCTATTTATGATTTCTATAAGCTATTTTTCAGCAGAAATGAGATATTAAACCCAATGCTTTG	969
Db	508	CAGCTCTTTATGATTTCCATAAGTGATTTTCAGCCGAATGAGGTATTACACACA-----TTG	454
Qy	970	ATTGTTCTAGAAAGTATAGTAATTTGTTTCTTAAGTGGTTCAAGCA-TCTACTCTTTT	1028
Db	453	ATCGTTCTAGAAGATACAGTAATTTGTTTCTTAAAGGGTCCACACATTTTTCCTTTT	394
Qy	1029	ATCATTACTTCAAAAATGACATTTGCTAAAGACTGCATTATTTTACTACTGTAATTTCTCC	1088
Db	393	ATCAGTTACTTCAAAAATGACATTTGTTGA-AGACAATTATTTTACAACCGTGTTTTCTCT	335
Qy	1089	ACGACATAGCATTATGTACATAGATGAGTGTAAACATTTATATCTCAC----ATAGAGACA	1144
Db	334	GTGACATAGCGTTATGTATATAGATGAGAGTGACGTTTCTATCTCACATAAATAGAGATG	275
Qy	1145	TGCTTATATGTTTATTATTAATAATGAAATGCCAGTCCATTCACACTGACATAAATAGAACTC	1204
Db	274	GGCTTAGATGTTTCTATTTTAAATGAAATACTGATTCATTACACTGAATAAATAGAAATTC	215
Qy	1205	AACTATTGCTTTTCAGGGAAATCATGATGATAGGTTGAAGAAGGTTACTATTATTAATTGTTTA	1264
Db	214	AACTATTGCTTTTCAGGGGACCGGGGATAAGATTGAAGAAGGTTAATAATTGTTTA	155
Qy	1265	AAAACAGCTTAGCGATTAAATGTCCTCCATTTTATATGAAGATTAAATGAAGGCTTTTAAT	1324
Db	154	AAAACGGCTTAGTGATGAATGCACCTTGATTTTATATGAAGGTTAAATGAAGGCTTTTAAT	95
Qy	1325	CAGCATTGTAAAGGAAATTAAGATGCTTTTCTGATATGCTGTTTTTTAGCCCTAGGAGTTAG	1384
Db	94	CAGCAGTGTAAAGGAACTAAATGGCTTTCTGATATCTCTGTTTTTTCAGCCTAGGAGTTAG	35
Qy	1385	AAATCCTAACTCTTTTATCCTCTCTCC	1413
Db	34	AAATCCTAACTCTTTTATCCTCTCTCC	6
RESULT 25			
CF169015			
LOCUS			
DEFINITION	CF169015	693 bp mRNA linear	EST 25-JUL-2003
	B0808A07-5	NIA Mouse Newborn Kidney cDNA Library (Long 1)	Mus
	musculus	cDNA clone NIA:B0808A07	IMAGE:30468198 5', mRNA sequence.
ACCESSION	CF169015		
VERSION	CF169015.1	GI:33278564	
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
	1 (bases 1 to 693)		
REFERENCE	Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.		
AUTHORS	Construction of long-transcript enriched cDNA libraries from		
TITLE	submicrogram amounts of total RNAs by a universal PCR amplification		
	method		
JOURNAL	Genome Res. 11 (9), 1553-1558 (2001)		
MEDLINE	21429098		
PUBMED	11544199		
COMMENT	Contact: Dawood B. Dudekula		
	Laboratory of Genetics		
	National Institute on Aging/National Institutes of Health		
	333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA		
	Email: cdna@igsun.grc.nia.nih.gov		
	Plate: B0808 row: A column: 07		
	Seg primer: M13 Reverse		
	High quality sequence stop: 693		
	POLYA=No.		
FEATURES	Location/Qualifiers		

	source	1. .693	
		/organism="Mus musculus"	
		/mol_type="mRNA"	
		/strain="C57BL/6J"	
		/db_xref="niaEST:B0808A07-5"	
		/db_xref="taxon:10090"	
		/clone="NIA:B0808A07 IMAGE:30468198"	
		/dev_stage="Newborn Kidney"	
		/lab_host="DH10B"	
		/clone_lib="NIA Mouse Newborn Kidney cDNA Library (Long 1)"	
		/note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: Sali; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). In brief, double-stranded cDNAs were synthesized with an oligo(dT) primer (Invitrogen): 5'-pGACTAGTTCTAGATCGGAGCGCGCCCTTTT-3' from 26 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with Sali and NotI enzymes and cloned into Sali/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3.0 kb. The library was constructed by Yulan Piao."	
ORIGIN			
	Query Match	21.3%;	Score 427.8; DB 14; Length 693;
	Best Local Similarity	83.4%;	Pred. No. 2.2e-75;
	Matches	486; Conservative	0; Mismatches 97; Indels 0; Gaps 0;
Qy	90	GAGGATAATGGCAACCCATGCCCTTAGAAAAATCGTGGGTGTTCTTGGTGGTGTGGAAT	149
Db	111	GAGGACGATGGCAACCTACGCTCTTCAAATGGCTGCACTGGTGTGGTGTGGCAT	170
Qy	150	GCTGGGCACAGTGGTGTCTACTGTTCATGCCTCAGTGGAGAGTGTGGGCTTCATTGAAAA	209
Db	171	GCTGGGCACCGTGGTGTGACTATCATGCCTCAGTGGAGAGTGTCTGCCTTCATCGAAAG	230
Qy	210	CAACATCGTGTGTTTGAATACTTCTGGGAAGGACTGTGGATGAATTCGTGAGGCAGGC	269
Db	231	TAACATTGTGTTGTGAGAACCGCTGGGAAGGCTTGTGGATGAATTTGATGAGGCATGC	290
Qy	270	TAAATCAGGATGAGTGCAGTGCAGAAATCTATGATTCCTGCTGGCTCTTTCTCCGACCTACA	329
Db	291	CAACATCAGAAATGAGTGCAGTGCAGTGCAGTGCCTGCTGGTCTTTAGTCCAGACCTCCA	350
Qy	330	GGCAGCCAGAGGACTGATGTGTGCTGCTTCCGTGATGTCCTTCTTGGCTTTTCATGATGC	389
Db	351	GGCATCCCGAGGACTGATGTGTGCTGCTGCTGCTTCTTGGCTTTTATGACAGC	410
Qy	390	CATCCTTGGCATGAATGCACCAGGTGCACGGGGACAAATGAGAAGGTGAAGGCTCACAT	449
Db	411	CATCCTCGGAATGAAGTGCACCAGATGCACGGGGACGATGAGAACGTGAAGAGCCGCAT	470
Qy	450	TCTGCTGACCGCTGGAATCATCTTTCATCATCACGGGGCATGGTGGTGTCTATCCCTGTGAG	509
Db	471	CTTGCTGACAGCCCGAATCATCTTCTTCATCATCACCGGCTGGTGTGTCTATCCCTGTGAG	530
Qy	510	CTGGGTTGCCAATGCCATCATCAGAGATTTCTATCACTCAATAGTAGTAATGTTGCCAAAA	569
Db	531	CTGGGTTGCCAATTCATCATCAGAGACTTCTACACCCACTGGTGGATGTGGCCCTAAA	590
Qy	570	ACGTGAGCTTGGAGAAGCTCTCTACTTAGGATGGACACGGCAGCTGGTGTGATTGTTGG	629
Db	591	GCGGAGCTGGGAGAAAGCCCTCTACATAGGCTGGTGGACACAGCGCTGGTGTGATCGCTGG	650

QY 630 AGGAGCTCTGTTCTGCGGTTTTTTTGTGCAACGAAAGAGC 672
|||||
Db 651 AGGAGCACTGTTCTGTGTGTGTTTGTGTACTGAAAGGAGC 693
|||||

RESULT 26
AA536178/c 490 bp mRNA linear EST 21-AUG-1997
LOCUS nf96b03.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:927725 3',
mRNA sequence.
ACCESSION AA536178
VERSION AA536178.1 GI:2280431
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 490)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 572 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 315.
FEATURES Location/Qualifiers
1. .490
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:927725"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Co3"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from 12 pooled bulk tumor samples and primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library went through one round of
normalization."
ORIGIN
Query Match 21.1%; Score 425; DB 9; Length 490;
Best Local Similarity 100.0%; Pred. No. 8.6e-75;
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1586 TTGAAGAAGATGATGCATTTTGACAAGAAATCATATATGTATGGATATATTTTAAATAAG 1645
|||||
Db 425 TTGAAGAAGATGATGCATTTTGACAAGAAATCATATATGTATGGATATATTTTAAATAAG 366
|||||
QY 1646 TATTGAGTACAGACTTTGAGGTTTCATCAATATAATAAAAGACGAGAAATAATGTCT 1705
|||||
Db 365 TATTGAGTACAGACTTTGAGGTTTCATCAATATAATAAAAGACGAGAAATAATGTCT 306
|||||
QY 1706 TGGTTTTCATTTGCTTACCAAAAAACACACAAAAAAGTTGCTCTTTGAGAACTTCA 1765
|||||
Db 305 TGGTTTTCATTTGCTTACCAAAAAACACACAAAAAAGTTGCTCTTTGAGAACTTCA 246
|||||
QY 1766 CCTGCTCTATGTGGGTACTGAGTCAAATTTGTCATTTTGTGTAATAATAATT 1825
|||||

Db 245 CCTGCTCCTATGTGGGTACCTGAGTCAAAATTTGTCATTTTGTCTGAAAAATAAATT 186
|||||
QY 1826 TCCTTCTTGTACCATTTCTGTTTAGTTTTTACTAAAAATCTGTAAATACGTATTTTCTGT 1885
|||||
Db 185 TCCTTCTTGTACCATTTCTGTTTAGTTTTTACTAAAAATCTGTAAATACGTATTTTCTGT 126
|||||
QY 1886 TTATTCCAAATTTGATGAAACTGACAATCCAATTTGAAAAGTTTGTGCGACGTCTGTCTA 1945
|||||
Db 125 TTATTCCAAATTTGATGAAACTGACAATCCAATTTGAAAAGTTTGTGCGACGTCTGTCTA 66
|||||
QY 1946 GCTTAAATGAATGTGTTCTATTGCTTTTATACATTATATTAATAAATGTACATTTTTC 2005
|||||
Db 65 GCTTAAATGAATGTGTTCTATTGCTTTTATACATTATATTAATAAATGTACATTTTTC 6
|||||
QY 2006 TAATT 2010
|||||
Db 5 TAATT 1
|||||

RESULT 27
CF140548
LOCUS UI-HF-CB0-apk-b-01-0-UI.r2 NIH_MGC_210 Homo sapiens cDNA clone
DEFINITION IMAGE:3098017 5', mRNA sequence.
ACCESSION CF140548
VERSION CF140548.1 GI:33255992
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 444)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Tim Ratliff
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/humanfl.html
Seq primer: pYX-5.
FEATURES Location/Qualifiers
1. .444
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3098017"
/tissue_type="CNCAP(3)T-225 cell line"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_MGC_210"
/note="Organ: Prostate; Vector: pT7T3 Pac; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with Not I and then cloned
directionally into pT7T3 Pac vector. The library tag
sequence located between the Not I site and the polyA tail
is CCCAC. Tissue was provided by Tim Ratliff."
ORIGIN

Seq primer: -40UP from Gibco
High quality sequence stop: 379.
Location/Qualifiers
1. .400
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2690628"
/lab_host="DH10B"
/clone_lib="NCI CGAP_Kid11"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI CGAP Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaudo."

ORIGIN

Query Match 19.9%; Score 399; DB 10; Length 400;
Best Local Similarity 100.0%; Pred. No. 1.4e-69;
Matches 399; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1612 AGAAATCATATATGATGGATATATTTAATAAGTATTTGAGTACAGACTTTGAGGTTTC 1671
Db 400 AGAAATCATATATGATGGATATATTTAATAAGTATTTGAGTACAGACTTTGAGGTTTC 341
QY 1672 ATCAATATAAATAAGAGCAGAGAAAATATGTCCTGGTTTTCATTGCTTACCAAAAAA 1731
Db 340 ATCAATATAAATAAGAGCAGAGAAAATATGTCCTGGTTTTCATTGCTTACCAAAAAA 281
QY 1732 CAACAACAAAAAGTTGTCCTTTGAGAACTTCACCTGCTCCTATGTGGTACCTGAGTC 1791
Db 280 CAACAACAAAAAGTTGTCCTTTGAGAACTTCACCTGCTCCTATGTGGTACCTGAGTC 221
QY 1792 AAAATTGTCATTTTGTTCGTGAAAAATAAATTCCTTCTTGTAACCATTTCTGTTAGT 1851
Db 220 AAAATTGTCATTTTGTTCGTGAAAAATAAATTCCTTCTTGTAACCATTTCTGTTAGT 161
QY 1852 TTTACTAAATCTGTAAATCTGTAATATTTTCTGTTTATTCCTTCTTGTAACCATTTCTGTTAGT 1911
Db 160 TTTACTAAATCTGTAAATCTGTAATATTTTCTGTTTATTCCTTCTTGTAACCATTTCTGTTAGT 101
QY 1912 ATCCAAATTGAAAGTTTGTGTCGACGTCTGTAGCTTAAATGAATGTGTTCTATTGCT 1971
Db 100 ATCCAAATTGAAAGTTTGTGTCGACGTCTGTAGCTTAAATGAATGTGTTCTATTGCT 41
QY 1972 TTATACATTTATATTAATAAATTGTACATTTTCTAATT 2010
Db 40 TTATACATTTATATTAATAAATTGTACATTTTCTAATT 2

RESULT 30
CK333041 642 bp mRNA linear EST 19-DEC-2003
LOCUS H8229C06-5 NIA Mouse Unique Gene Set Version 2 Mus musculus cDNA
DEFINITION clone H8229C06 5', mRNA sequence.
ACCESSION CK333041
VERSION CK333041.1 GI:40232656
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 642)
AUTHORS VanBuren,V., Piao,Y., Dudekula,D.B., Qian,Y., Carter,M.G.,
Martin,P.R., Stagg,C.A., Bassey,U., Aiba,K., Hamatani,T.,
Kargul,G.J., Luo,A.G., Kelso,J., Hide,W. and Ko,M.S.H.
TITLE Assembly, verification, and initial annotation of NIA 7.4K mouse
cDNA clone set

JOURNAL Genome Res. 12 (12), 1999-2003 (2002)
MEDLINE 22354164
PUBMED 12466305
COMMENT Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: H8229 row: C column: 06
Seg primer: M13 Reverse
High quality sequence stop: 642
POLYA=No.

FEATURES

source Location/Qualifiers
1. .642
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="niaEST:H8229C06-5"
/db_xref="taxon:10090"
/clone="H8229C06"
/sex="mixed"
/dev_stage="mixed"
/lab_host="DH10B"
/clone_lib="NIA Mouse Unique Gene Set Version 2"
/note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This clone is among a rearranged set of 11,424 clones from more than 20 cDNA libraries."

ORIGIN

Query Match 19.8%; Score 398.6; DB 14; Length 642;
Best Local Similarity 83.5%; Pred. No. 1.6e-69;
Matches 452; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
QY 90 GAGGATAATGGCAACCCATGCCTTAGAAATCGCTGGCTGTTTCTTGGTGGTGTGGAAT 149
Db 102 GAGGACGATGGCAACCTACGCTCTTCAATGGCTGCACTGGTGTCTTGGTGGTGTGGAAT 161
QY 150 CGTGGGCACAGTGGCTGTCACTGTCAATGCCTCAGTGGAGAGTGTGCGCCTTCATTGAAAA 209
Db 162 GGTGGGCACGGTGGCTGTGACTATATGCCTCAGTGGAGAGTGTCTGCCCTTCATCGAAAG 221
QY 210 CAACATCGTGGTTTGTGAAAACTTCTGGGAAGGACTGTGGATGAATTCGTCGAGGCAGGC 269
Db 222 TAACATTGTGGTGTGAGAACCGCTGGGAAGGCTTGTGGATGAATCGTATGAGGCATGC 281
QY 270 TAACATCAGGATCGAGTGCAGAAATCTATGATTCCTCGTGGCTCTTTCTCCGACCTACA 329
Db 282 CAACATCAGAAATGCAGTGCAGAGTCTACGACTCCCTGCTGGCTCTTAGTCCAGACCTCCA 341
QY 330 GGCAGCCAGAGGACTGATGTGTGCTTCCGTTCGTGATGTCCTTCTTGGCTTTTCATGATGGC 389
Db 342 GGCATCCCAGGACTGATGTGTGCTGCGTCCGTCTTGGCTTTTTCATGACAGC 401
QY 390 CATCCTTGGCATGAAATGCACCAGTGCACGGGGACAAATGAGAAAGGTGAAGGCTCACAT 449
Db 402 CATCCTCGGAATGAAGTGCACCAGATGCACGGGGACGATGAGAAACGTGAAGAGCCGCAT 461
QY 450 TCTGCTGACGGCTGGAATCATCTTTCATCATCACGGGCATGGTGGTGTCTCATCCCTGTGAG 509
Db 462 CTTGCTGACAGCCGGAATCATCTTCTTCATCACCCGCTTGGTGTGTCTCATCCCTGTGAG 521
QY 510 CTGGGTTGCCAATGCCATCATCAGAGATTTCTATAACTCAATAGTGAATGTTGGCCCAAAA 569
Db 522 CTGGGTTGCCAATTCATCATCATCAGAGACTTCTACAACCCACTGGTGGATGTGGCCCTAAA 581
QY 570 ACGTGAGCTTGGAGAGCTCTCTACTTAGGATGGACCAACGACCTGGTGTGATGTTGTTGG 629
Db 582 GCGGAGCTGGGAGAGCCCTCTACATAGGCTGGACCACAGCGCTGGTGTGATCGCTGG 641
QY 630 A 630

Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:977992
Seq primer: custom primer used
High quality sequence stop: 518.
Location/Qualifiers
1. .551
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1921700"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Sugano mouse kidney mkia"
/note="Organ: kidney; Vector: pME18S-FL3; Site 1: DraIII (CACTGTGTG); Site 2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA."

ORIGIN
Query Match 19.7%; Score 396.6; DB 9; Length 551;
Best Local Similarity 83.5%; Pred. No. 4.1e-69;
Matches 450; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
QY 90 GAGGATAATGGCAACCCATGCTTAGAAATCGCTGGGCTGTTTCTTGGTGGTGGTGAAT 149
Db 13 GAGGACGATGGCAACCTACGCTCTTCAAATGSGTGCACTGGTGTGTTGGTGGCAT 72
QY 150 GGTGGGACAGTGGCTGTCACTGTCTATGCCCTCAGTGGAGAGTGGCGCTTCATTGAAAA 209
Db 73 GGTGGGACGGTGGCTGTGACTATCATATGCCTCAGTGGAGAGTGTGCCTTCATCGGAAG 132
QY 210 CAACATCGTGGTGTGTTTGAACACTTCTGGGAAGGACTGTGGATGAATTCGCTGAGGCAGGC 269
Db 133 TAACATGTGGTGTGTTGAGAACCGCTGGGAAGGCTTGTGGATGAATGTATGAGGCATGC 192
QY 270 TAACATCAGGATGCAGTGCAGAAATCTATGATCCCTGCTGGCTCTTCTCCGGACCTACA 329
Db 193 CAACATCAGAAATGCAGTGCAGAGGTCTACGACTCCCTGCTGGCTCTTAGTCCAGACCTCCA 252
QY 330 GGCAGCCAGAGGACTGATGTGTGCTGCTTCCGTGATGTCCTTCTGGCTTTTCATGATGC 389
Db 253 GGCATCCGAGGACTGATGTGTGCTGCTCCGTCCGTCTTGGCTTTTCATGACAGC 312
QY 390 CATCCTTGGCATGAATGCACAGGTGCACGGGGACAATGAGAGGTGAAGGCTCACAT 449
Db 313 CATCCTCGAATGAAGTGCACAGATGCACGGGGACGATGAGAACGTGAAGAGCCGCAT 372
QY 450 TCTGTGACGGCTGGAATCATCTTTCATCATCAGGGCATGGTGGTCTCATCCCTGTGAG 509
Db 373 CTTGTGACAGCCGGAATCATCTTCTCATCAGGCTTGGTGTGCTCATCCCTGTGAG 432
QY 510 CTGGGTTGCCAATGCCATCATCAGAGATTTCTATAACTCAATAGTGAATGTTGCCAAAAA 569
Db 433 CTGGGTTGCCAATTCATCATCAGAGACTTCTACAACCCACTGGTGGATGTGGCCCTAAA 492
QY 570 ACGTGAGCTTGAGAGCTCTCTACTTAGGATGGACCACGGCACTGGTGTGATGTTG 628
Db 493 GCGCGAGTGGGAGAAGCCCTCTACATAGGCTGGACCACAGCGCTGGTGTGATCGCTG 551

RESULT 33
Bi104115
LOCUS
DEFINITION 602890432F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5035638 5', mRNA sequence.
ACCESSION Bi104115
VERSION Bi104115
KEYWORDS EST.
SOURCE Bi104115.1 GI:14555008
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 754)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM11098 row: p column: 07
High quality sequence start: 2
High quality sequence stop: 653.
High quality sequence stop: 653.
Location/Qualifiers
1. .754
/organism="Mus musculus"
/mol_type="mRNA"
/strain="Czech II"
/db_xref="taxon:10090"
/clone="IMAGE:5035638"
/tissue_type="spontaneous tumor, metastatic to mammary."
Stem cell origin."
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Lu29"
/note="Organ: lung; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

ORIGIN
Query Match 19.4%; Score 390.8; DB 12; Length 754;
Best Local Similarity 78.3%; Pred. No. 5.5e-68;
Matches 506; Conservative 0; Mismatches 137; Indels 3; Gaps 3;
QY 90 GAGGATAATGGCAACCCATGCTTAGAAATCGCTGGGCTGTTTCTTGGTGGTGGTGAAT 149
Db 106 GAGGACCATGGCAACCTACGCTCTTCAAATGGCTGCACTGGTGTGTTGGTGGTGGCAT 165
QY 150 GGTGGGCACAGTGGC-TGTCACTGTCAATGCCCTCAG-TGGAGAGTGTCCGCTTCATTGAA 207
Db 166 GGTGGGCACGGTGGCTTGTGACTATATATGCCTCAGTTGGAGAGTGTCTGCCTTCATCGAA 225
QY 208 AACAAACATCGTGGTGTGTTTGAACACTTCTGGGAAGGACTGTGGATGAATTGCGTGAGGCAG 267
Db 226 AGTAACATTTGTGGTGTGTTGAGAACCGCTGGGAAGGCTTGTGGATGAACATGTATGAGGCAT 285
QY 268 GCTAACATCAGGATGCAGTGCAGAAATCTATGATTCCTGCTGGCTCTTTCTCCGGACCTA 327
Db 286 GCCAACATCAGAAATGCAGTGCAGAGTCTACGACTCCCTGTGGCTCTTAGTCCAGACCTC 345
QY 328 CAGGCAGCCAGAGGACTGATGTGTGCTGCTCCGTGATGTCCTTCTTGGCTTTCATGATG 387
Db 346 CAGGCATCCAG-GGACTGATGTGTGCTGCGTCCGTCTTGGCTTTCATGACA 404
QY 388 GCCATCCTTGGCATGAATGCACAGGTGCACGGGGACAATGAGAGGTGAAGGCTCAC 447
Db 405 GCCATCCTCGGAATGAAGTGCACAGATGCACGGGGGACCATGAGAACGTGAAGAGCCGC 464

Library made with RNA pooled from multiple tissues including liver, lung, hypothalamus, pituitary, and placenta/endometrium."

ORIGIN

Query Match 18.0%; Score 360.8; DB 14; Length 584;
Best Local Similarity 81.7%; Pred. No. 5.9e-62;
Matches 478; Conservative 0; Mismatches 98; Indels 9; Gaps 5;

QY 934 ATTTCAGCAGAAATGAGATATT-AAACCCAAATGCTTTGATTGTTCTAGAAAGTATAGTAAT 992
|||||
Db 1 ATTTCAGCTGCATGAGAAATGTAGCACATGCTTTGATTATTCTAAAAGAACAGTAAT 60

QY 993 TTGTTTCTAAGGTGTTCAAGCATCTACTCTTTTATCATTTACTTCAAAATGACATTG 1052
|||||
Db 61 TTGTTTCTAAAATGGGTCATGTTTATCTCTTTTATTAGTACTTCAAAATGACATTG 120

QY 1053 CTAAGACTGCATTATTTACTACTGTAATTCTCCAGCATAGCATTATGATACATAGA 1112
|||||
Db 121 TGGGAGAC--AATTATTTACAACCTGTGACTTCTCTATGACAGAGCGTTATGATATAGA 178

QY 1113 TGAGTGTAAACATTTATATCTCACA---TAGAGACATGCTTATATGTTTATTTAAAT 1168
|||||
Db 179 TGAGTGTGACATTTCTGTGTACATAGGTAGAGACAGGCTTATATAGTCTCTATTTAAAT 238

QY 1169 GAATGCCAGTCCATTACACTGAATAAATAAAGAACTCAACTATTGCTTTTCAGGGAATCA 1228
|||||
Db 239 GAACACTGATTCATTACACTGAATAAATAAAGAACTCAACTATTGCTTTTCAGGGAATCA 298

QY 1229 TGGATAGGTTGAAGAGGTTACTATTAAATGTTTAAACACAGCTTAGGGATTAAATGTC 1288
|||||
Db 299 GGGTAAGATTGAAGAGGTTAATAATTGTTTAAACACAGCTTAGGTATGATGATGCAC 358

QY 1289 TCCATTTATAATGAAGAT-TAAATGAAGGCTTTAATCAGCATTTGTAAGGAATGAAT 1347
|||||
Db 359 TTAATTTATAATGAAGGTTAATAATGGAAGCTTTAATCAGCAGGTTAAGGGAATGAAT 418

QY 1348 GGCCTTCTGATATGCTGTTTTCAGCTAGGAGTTAGAAATCCTAACTTCTTTATCCTCT 1407
|||||
Db 419 GGCCTTTCGATATCTCGTTTTCAGCCTTANGAGTTAGAAATCCAAATTCCTTTTCCTCA 478

QY 1408 TCTCCAGAGGCTTTTTCCTTCTGTTATTAATAATTAACAT-TTTTAAACGACATATT 1466
|||||
Db 479 TTCCCCAGAGGCTTCCCTTCTGTTATTAATAATGGGCATCTTTTAAAGACGATATT 538

QY 1467 TTGTCAAGGGGCTTTTGATTCAAACTGCTTTTCCAGGGCTATACT 1511
|||||
Db 539 TTGTCAAGGGGCTTTTGATTCAAACTGCTTTTGCAGAGCTATACT 583

RESULT 36
BE285566
LOCUS 601095576F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3490218 5',
DEFINITION mRNA sequence.
ACCESSION BE285566
VERSION BE285566.1 GI:9163274
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 918)
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM8532 row: 9 column: 19
High quality sequence stop: 629.

FEATURES

source
Location/Qualifiers
1..918
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3490218"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam5"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

ORIGIN

Query Match 17.9%; Score 360.4; DB 10; Length 918;
Best Local Similarity 83.0%; Pred. No. 6.5e-62;
Matches 470; Conservative 0; Mismatches 91; Indels 5; Gaps 5;

QY 90 GAGGATAATGGCAACCCATGCTTAGAAATCGCTGGGCTGTTCTTTGGTGGTGAAT 149
|||||
Db 74 GAGGACGATGGCAACCTACGCTCTTCAAATGGCTGCACTGGTCTTGGTGGTGGCAT 133

QY 150 GGTGGGCACAGTGGCTGTCACTGTTCATGCCTCAGTGGAGAGTGTGGGCTTCATTGAAA 209
|||||
Db 134 GGTGGGCACGTTGGCTGTGACTATCATGCCTCAGTGGAGAGTGTCTGCCTTCATCGAAAG 193

QY 210 CAACATCGTGGTTTTGAAACCTTCTGGGAAGGACTGTGGATGAATTGCGTGAGGCAGGC 269
|||||
Db 194 TAAACATTGTGGTGTGAGAACCGCTGGGAAGGCTGTGGATGAATTGTATGAGGCATGC 253

QY 270 TAAC-ATCAGGATGCAGTG-CAAAATCTATGATTCCTGCTGGCTCTTCTCCGGACCTA 327
|||||
Db 254 CAACGATCAGAATGCAGTGCCAAAGGTCTACGACTCCCTGCTGGCTTCTAGTCCAGACCTC 313

QY 328 CAGGCAGCCAGAGGACTGATGTGTGCTGCTTCCGTGATGTCCTTCTGGCTTTCATGATG 387
|||||
Db 314 CAGGCATCCCGAGGACTGATGTGTGCTGCTCCGTCCGTCTTCTGGCTTTCATGACG 373

QY 388 -GCCATCCTTGGCATGAAATGCACAGGTGCACGGGGGACAATGAGAAGGTGAAGGCTCA 446
|||||
Db 374 AGCCATCCTCGAATGAAGTGACACAGATGCACGGGGACGATGAGAACGTGAAGAGCCG 433

QY 447 CATTTCTGCTGACGGCTGGAATCATCTTTCATCATCACGGGCGCATGGTGGTCTCATCCCTGT 506
|||||
Db 434 CATCTTGTGACAGCCGGAATCATCTTCTTCATCACCGGCTTGGTGTGTGCTCATCCCTGT 493

QY 507 GAGCTGGGTGCCAATGCCATCATCAGAGATTTCTATACTCAATAGTGAATGTTGCCCA 566
|||||
Db 494 CAGCTGGGTGCCAATTCATCATCAGAGACTTCTACACCCACTGGTGGATGTGGCCCT 553

QY 567 AAAACGTGAGCTTGGAGAGCTTCTACTTAGGATGACCCACGGCACCTGGTCTGATTGT 626
|||||
Db 554 AAAGCGTGAGCTGGGAGAGC-CTCTACATAGGCTGGACCAACAGCGCT-GTGTGATCGC 611

QY 627 TGGAGGAGCTCTGTTCTGCTGCGTTT 652
|||||
Db 612 TGGATGAGCCCTGTTCTGTGTGTGT 637

RESULT 37
CB235084
LOCUS
DEFINITION
ACCESSION

CB235084
AGENCOURT 11444483 NIH MGC 166 Mus musculus cDNA clone
IMAGE:30246099 5', mRNA sequence.

767 bp
mRNA linear
EST 10-FEB-2003

Db 1 ACAGAGAACCCCTGCTTCAAAGCAGAAAGTAGCAGTTCGGAGTCCAGCTGGCTAAAACTCA 60

QY 85 TCCCAGAGGATAATGSCAACCCCATGCCCTTAGAAATCGCTGGGCTGTTTCTTGGTGTGTT 144

Db 61 TCCCAGAGGATAATGGCAACCCCATGCCCTTAGAAATCGCTGGGCTGTTTCTTGGTGTGTT 120

QY 145 GGAATGGTGGGCACAGTGGCTGTCACTGTCTATGCCCTCAGTGGAGAGTGTGCGGCCTTCATT 204

Db 121 GGAATGGTGGGCACAGTGGCTGTCACTGTCTATGCCCTCAGTGGAGAGTGTGCGGCCTTCATT 180

QY 205 GAAACAACATCGTGGTTTTTGAACCTTCTGGGAAGGACTGTGGATGAATTGCGTGAGG 264

Db 181 GAAACAACATCGTGGTTTTTGAACCTTCTGGGAAGGACTGTGGATGAATTGCGTGAGG 240

QY 265 CAGGCTAACATCAGGATGCAGTGCAGAAATCTATGATTCCTCGTGGCTCTTTTCTCGGAC 324

Db 241 CAGGCTAACATCAGGATGCAGTGCAGAAATCTATGATTCCTCGTGGCTCTTTTCTCGGAC 300

QY 325 CTACAGGCAGCCAGAGGACTGATGTGTGCTGCTTCCGTGATGTCCTTCTTGGCTT 379

Db 301 CTACAGGCAGCCAGAGGACTGATGTGTGCTGCTTCCGTGATGTCCTTCTTGGCTT 355

RESULT 39

BE172635

LOCUS BE172635 360 bp mRNA linear EST 21-JUN-2000

DEFINITION MR0-HT0559-050400-012-e04 HT0559 Homo sapiens cDNA, mRNA sequence.

ACCESSION BE172635

VERSION BE172635.1 GI:8635361

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

PUBMED 10737800

COMMENT Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR0-HT0559-050400-012-e04&t3=2000-04-05&t4=1) Seq primer: puc 18 forward High quality sequence start: 15 High quality sequence stop: 360. Location/Qualifiers 1. .360

FEATURES

source

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/dev_stage="Adult"

/clone_lib="HT0559"

/note="Organ: head neck; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 17.5%; Score 351; DB 10; Length 360;

Best Local Similarity 98.6%; Pred. No. 6.1e-60;

Matches 354; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1082 TTTCTCCACGACATAGCATTATGTACATAGATGAGTGTAAACATTTATATCTCACATAGAG 1141

Db 1 TTTCTACACGGCATAGCATTATGTACATAGATGAGTGTAAACATTTATATCTCACATAGAG 60

QY 1142 ACATGCTTATATGGTTTTTATTAAATGAAATGCCAGTCCATTACACTGAATAATAGAA 1201

Db 61 ACATGCTTATATGGTTTTTATTAAATGAAATGCCAGTCCATTACACTGAATAATAGAA 120

QY 1202 CTCAACTATTGCTTTTCAGGGAAATCATGGATAGGTTGAAGAAGTTACTATTAAATTGT 1261

Db 121 CTCAACTATTGCTTTTCAGGGAAATCATGGATAGGTTGAAGAAGTTACTATTAAATTGT 180

QY 1262 TTAAAAACAGCTTAGGGATTAATGTCTCCATTTATAATGAAGATTAAATGAAGGCTTTT 1321

Db 181 TTAAAAACAGCTTAGGGATTAATGTCTCCATTTATAATGAAGATTAAATGAAGGCTTTT 240

QY 1322 AATCAGCATTGTAAAGGAAATCAATGGCTTCTCATATGCTGTTTTTTAGCCTAGGAGT 1381

Db 241 AATCAGCATTGTAAAGGAAATCAATGGCTTCTCATATGCTGTTTTTTAGCCTAGGAGT 300

QY 1382 TAGAAATCCTAACTTCTTTATCCTCTTCTCCAGAGGCTTTTTTTTTCTGTATTAA 1440

Db 301 TAGAAATCCTAACTTCTTTATCCTCTTCTCCAGAGGCTTTTTTTTTCTGTAGAAA 359

RESULT 40

BE000282

LOCUS BE000282 352 bp mRNA linear EST 05-JUN-2000

DEFINITION MR0-BN0070-260400-017-e10 BN0070 Homo sapiens cDNA, mRNA sequence.

ACCESSION BE000282

VERSION BE000282.1 GI:8260515

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 352) Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

PUBMED 10737800

COMMENT Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR0-BN0070-260400-017-e10&t3=2000-04-26&t4=1) Seq primer: puc 18 forward High quality sequence start: 7 High quality sequence stop: 352. Location/Qualifiers 1. .352

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="BN0070"
/note="Organ: breast normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN

Query Match 17.3%; Score 347.2; DB 10; Length 352;
Best Local Similarity 99.1%; Pred. No. 3.5e-59;
Matches 349; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1080 AATTCTCCAGCATAGCATTATGTACATAGTAGTGTAACATTATATCTCACATAG 1139
Db 1 ATTGTCTACAGCATAGCATTATGTACATAGTAGTGTAACATTATATCTCACATAG 60
QY 1140 AGACATGCTTATATGGTTTTTAAATGAATGCCAGTCCATTACACTGAATAATAG 1199
Db 61 AGACATGCTTATATGGTTTTTAAATGAATGCCAGTCCATTACACTGAATAATAG 120
QY 1200 AACTCAACTATTGCTTTTCAGGAAATCATGGATAGGGTTGAAGAAGGTTACTATTAA 1259
Db 121 AACTCAACTATTGCTTTTCAGGAAATCATGGATAGGGTTGAAGAAGGTTACTATTAA 180
QY 1260 GTTTAAAAACAGCTTAGGATTAAATGTCTCCATTATAATGAAGATTAAAGGCT 1319
Db 181 GTTTAAAAACAGCTTAGGATTAAATGTCTCCATTATAATGAAGATTAAAGGCT 240
QY 1320 TTAATCAGCAATTGTAAAGGAATTGAATGGCTTTCTGATATGCTGTTTTTAGCCTAGGA 1379
Db 241 TTAATCAGCAATTGTAAAGGAATTGAATGGCTTTCTGATATGCTGTTTTTAGCCTAGGA 300
QY 1380 GTTAGAAATCCTAACTTCTTTATCCTCTTCTCCAGAGGCTTTTTTTTCTT 1431
Db 301 GTTAGAAATCCTAACTTCTTTATCCTCTTCTCCAGAGGCTTTTTTTTCTT 352

RESULT 41
BE814313/c
LOCUS BE814313 351 bp mRNA linear EST 21-SEP-2000
DEFINITION MRO-BN0070-270500-026-e03 BN0070 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE814313
VERSION BE814313.1 GI:10246547
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 351)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2=MRO-BN0070-270
500-026-e03&t3=2000-05-27&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 351.
Location/Qualifiers
1. .351
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="BN0070"
/note="Organ: breast normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN

Query Match 17.2%; Score 345.2; DB 10; Length 351;
Best Local Similarity 99.1%; Pred. No. 8.8e-59;
Matches 347; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1091 GACATAGCATTATGTACATAGTAGTGTAACATTATATCTCACATAGACATGCTTA 1150
Db 351 GACATAGCATTATGTACATAGTAGTGTAACATTATATCTCACATAGACATGCTTA 292
QY 1151 TATGGTTTTTAAATGAATGCCAGTCCATTACACTGAATAAATAGAACTCAACTAT 1210
Db 291 TATGGTTTTTAAATGAATGCCAGTCCATTACACTGAATAAATAGAACTCAACTAT 232
QY 1211 TGCTTTTCAGGAAATCATGGATAGGGTTGAAGAAGGTTACTATTATTTTAAAAACA 1270
Db 231 TGCTTTTCAGGAAATCATGGATAGGGTTGAAGAAGGTTACTATTATTTTAAAAACA 172
QY 1271 GCTTAGGGATTAAATGTCTCCATTATAATGAAGATTAAATGAAGGCTTTAATCAGCAT 1330
Db 171 GCTTAGGGATTAAATGTCTCCATTATAATGAAGATTAAATGAAGGCTTTAATCAGCAT 112
QY 1331 TGTAAGGAAATTGAATGGCTTTCTGATATGCTGTTTTTAGCCTAGGAGTTAGAAATCC 1390
Db 111 TGTAAGGAAATTGAATGGCTTTCTGATATGCTGTTTTTAGCCTAGGAGTTAGAAATCC 52
QY 1391 TAACTTCTTTATCCTCTTCTCCAGAGGCTTTTTTTTCTTGTGTATTAA 1440
Db 51 TAACTTCTTTATCCTCTTCTCCAGAGGCTTTTTTTTCTTGTGTAGCAA 2

RESULT 42
BB618773
LOCUS BB618773 596 bp mRNA linear EST 26-OCT-2001
DEFINITION BB618773 RIKEN full-length enriched, 8 days embryo Mus musculus
cDNA clone 5730424M09 5', mRNA sequence.
ACCESSION BB618773
VERSION BB618773.1 GI:16458253
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 596)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.

TITLE
JOURNAL
COMMENT

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalizaton and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.

FEATURES
source

Location/Qualifiers
1. .596
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="5730424M09"
/sex="mixed"
/dev_stage="8 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 8 days embryo"
/note="Site 1: Sali; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTCTTN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of subtraction to
Rot = 100.0 Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATTCGAGTTAATAAATTAATCCCCCCCCCCC 3']. cDNA
was cloned into the XhoI and BamHI sites. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I. Cloning sites, 5' end: Sali; 3' end: BamHI."

ORIGIN

Query Match 16.9%; Score 339.8; DB 10; Length 596;
Best local Similarity 84.2%; Pred. No. 9.5e-58;
Matches 383; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 90 GAGGATAATGGCAACCCATGCCTTAGAAATCGCTGGCTGTTTCTTGGTGGTGGTGAAT 149
||||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| |||
Db 142 GAGGACGATGGCAACCTACGCTCTTCAAATGGCTGCACCTGGTGGTGGTGGTGGCAT 201
||||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| |||

QY 150 GGTGGGCACAGTGGCTGTCACTGTTCATGCCTCAGTGGAGAGTGTGGCCTTCATTGAAA 209
||||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| |||
Db 202 GGTGGGCACAGTGGCTGTCACTGTTCATGCCTCAGTGGAGAGTGTCTGCCTTCATCGAA 261
||||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| |||
QY 210 CAACATCGTGGTGTGTTTGAACAACTTCTGGGAAGGACTGTGGATGAATTGGCTGAGGCAG 269
||||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| |||
Db 262 TAACATTGTGGTGTGTTGAGAACCGCTGGGAAGGCTGTGGATGAATTGTATGAGGCATGC 321
||||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| |||
QY 270 TAACATCAGGATGCAGTGCACAAATCTATGATTCCTGCTGGCTCTTTCTCGGACCTACA 329
||||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| |||
Db 322 CAACATCAGAATGCAGTGCAGAGTCTACGACTCCCTGCTGGCTCTTAGTCCAGACCTCCA 381
||||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| |||
QY 330 GGCAGCCAGAGGACTGATGTGTGCTGCTTCCGTGATGTCTTCTTGGCTTTCATGATGC 389
||||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| |||
Db 382 GGCATCCCGAGGACTGATGTGTGCTGCTCCGTCTTGGCTTCTTCTTGGCTTTCATGACAGC 441
||||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| |||
QY 390 CATCCTTGGCATGAAATGCACCAAGTGCACGGGGGACAAATGAGAAAGGTGAAGGCTCACAT 449
||||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| |||
Db 442 CATCCTCGGAATGAAGTGCACCAAGTGCACGGGGGACGATGAGAACGTGAAGAGCCGCAT 501
||||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| |||
QY 450 TCTGCTGACGCTGGAATCATCTTCATCATCACGGGCGATGGTGGTGCATCCCTGTGAG 509
||||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| |||
Db 502 CTTGCTGACAGCCGGAATCATCTTTCATCACCCGCTTGGTGTGCTCATCCCTGTGAG 561
||||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| |||
QY 510 CTGGGTTGCCAATGCCATCATCAGAGATTTCTATA 544
||||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| |||
Db 562 CTGGGTTGCCAATTCATCATCAGAGACTTCTACA 596
||||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| |||

RESULT 43
CB306981/c
LOCUS
DEFINITION
CB306981
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
. PUBMED
COMMENT

CB306981 363 bp mRNA linear EST 04-MAR-2003
UI-CF-FN0-aen-j-20-0-UI.s1 UI-CF-FN0 Homo sapiens cDNA clone
UI-CF-FN0-aen-j-20-0-UI 3', mRNA sequence.
CB306981
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 363)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
The following repetitive elements were found in this cDNA
sequence: 1-58, >AT rich#Low_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
Location/Qualifiers
1. .363
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-FN0-aen-j-20-0-UI"

/tissue_type="Human Lung Epithelial cells"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-FN0"
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-FN0 is a subtracted cDNA library derived from two
normalized Human lung epithelial cell libraries (EN1 and
DUI1) The library was subtracted according to according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. For additional information, contact:
bento-soares@uiowa.edu
TAG_TISSUE=Lung Epithelial Cells Tissue nos 359-368
TAG_LIB=UI-CF-FN0
TAG_SEQ=GGCTGTAGGC"

ORIGIN

Query Match 16.7%; Score 336; DB 14; Length 363;
Best Local Similarity 100.0%; Pred. No. 6.1e-57;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1675 AATATAATAAAGAGCAGAAAAATATGTCTTGGTTTTCATTGCTTACCAAAAAACAA 1734
Db |||||||
QY 354 AATATAATAAAGAGCAGAAAAATATGTCTTGGTTTTCATTGCTTACCAAAAAACAA 295
Db |||||||
QY 1735 CAACAAAAAAGTTGTCCTTTGAGAACTTCACCTGCTCTATGTGGTACCTGAGTCAA 1794
Db |||||||
QY 294 CAACAAAAAAGTTGTCCTTTGAGAACTTCACCTGCTCTATGTGGTACCTGAGTCAA 235
Db |||||||
QY 1795 ATTGTCAATTTTGTCTGTGAAAAATAAATTCCTTCTGTACCATTTCTGTTAGTTT 1854
Db |||||||
QY 234 ATTGTCAATTTTGTCTGTGAAAAATAAATTCCTTCTGTACCATTTCTGTTAGTTT 175
Db |||||||
QY 1855 ACTAAAACTGTAAATACTGTATTTTCTGTTTATCCAAATTTGATGAAACTGACAATC 1914
Db |||||||
QY 174 ACTAAAACTGTAAATACTGTATTTTCTGTTTATCCAAATTTGATGAAACTGACAATC 115
Db |||||||
QY 1915 CAATTTGAAAGTTTGTGTCGACGCTGTCTAGCTTAAATGAATGTGTTCTATTGCTTTA 1974
Db |||||||
QY 114 CAATTTGAAAGTTTGTGTCGACGCTGTCTAGCTTAAATGAATGTGTTCTATTGCTTTA 55
Db |||||||
QY 1975 TACATTTATTAATAAATTTGTACATTTTCTAATT 2010
Db |||||||
QY 54 TACATTTATTAATAAATTTGTACATTTTCTAATT 19
Db |||||||

RESULT 44
CA339169
LOCUS
DEFINITION NISC lx11e07.y1 NCI CGAP Pr51 Rattus norvegicus cDNA clone
IMAGE:5621868 5', mRNA sequence.
ACCESSION CA339169
VERSION CA339169.1 GI:24557267
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 568)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
CDNA Library Preparation:
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov

FEATURES
source

Seq primer: M13RP1 reverse primer (ABI).
Location/Qualifiers
1..568
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:5621868"
/sex="male"
/tissue_type="pool of ventral and dorsolateral prostate"
/dev_stage="adult, 10 week"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Pr51"
/note="Organ: prostate; Vector: pCMV-SPORT6.1; Site 1:
NotI; Site 2: EcoRV; Cloned unidirectionally. Primer:
Oligo dt. Pool of 2 primary libraries: NCI CGAP Pr46
(ventral prostate from 10 wk male, average insert size 2
kb) and NCI CGAP Pr47 (dorsolateral prostate from 10 wk
male, average insert size 2 kb). Constructed by
Invitrogen. Note: this is a NCI CGAP Library."

ORIGIN

Query Match 16.5%; Score 332.4; DB 14; Length 568;
Best Local Similarity 79.5%; Pred. No. 2.9e-56;
Matches 404; Conservative 0; Mismatches 103; Indels 1; Gaps 1;
QY 35 CTGCTTCAAAGCAGAAAGTAGCAGTTCCGGAGTCCAGCTGGCTAAAACTCATCCAGAGGA 94
Db |||||||
QY 61 CTGCTTGAGAGCAGTCTACTTCTCCCTGTTTCAGCCTTCAGCTGGACGAAACTTGGGAAGA 120
Db |||||||
QY 95 TAATGGCAACCCATGCCTTAGAAATCGCTGGGCTGTTCTTGGTGGTGTGGAATGGTGG 154
Db |||||||
QY 121 CCATGGCAACCTACGCTCTTCAAATGGCTGCCCTGGTGGTGTGGCATGGTGG 180
Db |||||||
QY 155 GCACAGTGGCTGTCACTGTCTATGCTCAGTGGAGAGTGTGGCTTTCATTGAAACAACA 214
Db |||||||
QY 181 GCACAGTGGCTGTCACTGTCTATGCTCAGTGGAGAGTGTGGCTTTCATTGAAAGTAACA 240
Db |||||||
QY 215 TCGTGGTTTTTGAAAACTTCTGGGAGGAGCTGTGGATGAATTCGCTGAGGAGGCTAACA 274
Db |||||||
QY 241 TTGTCGTGTTTGAAACCGCTGGGAGGTTTGTGGATGAACCTGCATGAGGCGATGCCAACA 300
Db |||||||
QY 275 TCAGGATGCAGTGCAAAATCTATGATTCCTGTGGCTCTTTCTCCGACCTACAGGCAG 334
Db |||||||
QY 301 TCAGATGCAGTGAAGGTCTATGACTCCCTGTGGCTCTTAGTCCAGACCTCCNGGCAT 360
Db |||||||
QY 335 CCAGAGGACTGATGTGTCTGCTTCCGTGATGTCTTCTTGGCTTTTCATGATGCCCATCC 394
Db |||||||
QY 361 CCAGAGGACTGATGTGTCTGCTGCTGCTCCTGTCTTCTTGGCTTTTCATGACAGCATCC 420
Db |||||||
QY 395 TTGGCATGAAATGACCCAGGTGCACGGGGGACAATGAGAAGGTGAAGGCTCACATTCTGC 454
Db |||||||
QY 421 TCGAATGAAGTGCACAGGTGCACCGGGGAGGACGAGAACGTGAAGAGCGCATCTTGC 480
Db |||||||
QY 455 TGACGGCTGGAATCATCTTTCATCATCA-CGGGCATGGTGGTGTCTATCCCTGTGAGCTGG 513
Db |||||||
QY 481 TGACGGCTGGAATCATCTTCTTTCATCACCGGGCTTGGTGTGTCTATCCCTGTGAGCTGG 540
Db |||||||
QY 514 GTTGCCAATGCCATCATCAGAGATTTCCT 541
Db |||||||
QY 541 GTTGCCAATGCCATCATCAGAGATTTCCT 568
Db |||||||

RESULT 45
AI048929
LOCUS
DEFINITION uc84g08.y1 Sugano mouse kidney mKia Mus musculus cDNA clone
IMAGE:1432382 5' similar to TR:O35054 O35054 CPE-RECEPTOR. ;, mRNA
sequence.
ACCESSION AI048929
VERSION AI048929.1 GI:3297216
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

AI048929 465 bp mRNA linear EST 08-JUL-1998
uc84g08.y1 Sugano mouse kidney mKia Mus musculus cDNA clone
IMAGE:1432382 5' similar to TR:O35054 O35054 CPE-RECEPTOR. ;, mRNA
sequence.
AI048929
AI048929.1 GI:3297216
EST.
Mus musculus (house mouse)
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 465)

REFERENCE
AUTHORS
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE
JOURNAL
COMMENT
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicinep
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:916450

FEATURES
source
1. .465
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1432382"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/lab_lib="Sugano mouse kidney mkia"
/note="Organ: kidney; Vector: pME18S-FL3; Site 1: DraIII (CACTGTGTG); Site 2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTTAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA."

ORIGIN
Query Match 16.4%; Score 330.6; DB 9; Length 465;
Best Local Similarity 81.9%; Pred. No. 7e-56;
Matches 381; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 103 ACCCATGCTTAGAAATCGCTGGCTGTTTCTTGGTGGTGTGAATGGTGGCACAGTG 162
|||
1 ACCTACGCTCTTCAAAATGGCTGCTCTGGTGTGGTGGCTGGCATGGTGGCACGTTG 60
|||
163 GCTGTCAGTGTATGCCTCAGTGGAGAGTGTGCGCCTTCATTGAAAACAACATCGTGTT 222
|||
61 TCTGTTTCTATCATGCTCAGTGCACAGCGTCTGCCCTTCATCGAAAGTAACATTGTGTG 120
|||
223 TTTGAAACTTCTGGGAAGGACTGTGGATGAATTGCGTGAGGCAGGCTAACATCAGGATG 282
|||
121 TTTGAGAACCGCTGGGAAGGCTGTGGATGAATTGTATGAGGCATGCCAACATCAGAAATG 180
|||
283 CAGTGCAAAATCTATGATTCCTGCTGGCTCTTCTCCGGACCTACAGGCAGCCAGAGGA 342
|||
181 CAGTGCAAGGTCTACGACTCCCTGCTGGCTCTTAGTCCAGACCTCCAGGCATCCCGAGGA 240
|||
343 CTGATGTGTGCTGCTCCGTTGATGTGCTCTTCTTGGCTTTCATGATGCCATCCTTGGCATG 402
|||
241 CTGATGTGTGCTGCTCCGTTGCTGCTTCTTGGCTTTCATGACAGCCATCCTCGGAATG 300
|||
403 AAATGCACCAAGGTGCACGGGGGACAATGAGAAAGGTGAAGGCTCAATTCTCTGCTACGGCT 462

301 AAGTGCACCATGTCACGGGGACGATGAGAACTGAAGAGCCGCACTTGTGACAGCC 360
|||
463 GGAATCATCTTTCATCATCAGGGCATGGTGGTGTCTATCCCTGTGAGCTGGTGGCCAAT 522
|||
361 GGAATCATCTTCTTCATCACCAGCTTGGTGTGTGTCTATCCCTGTGAGCTGGTGGCCAAT 420
|||
523 GCCATCATCAGAGATTTCTATAAATACTCAATAGTGAATGTGCCCCAA 567
|||
421 TCCATCATCAGAGACTTCTACAACCCCACTGGTGGATGTGCCCTTA 465
|||

RESULT 46
AI547146 771 bp mRNA linear EST 09-AUG-1999
LOCUS PN3.1_01_G02.r mynorm Homo sapiens cDNA 5', mRNA sequence.
DEFINITION AI547146
ACCESSION AI547146
VERSION AI547146.1 GI:4464634
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 771)
Huang,G.M., Ng,W.L., Farkas,J., He,L., Liang,H.A., Gordon,D., Yu,J. and Hood,L.
Prostate cancer expression profiling by cDNA sequencing analysis
Genomics 59 (2), 178-186 (1999)
99339982
10409429
PUBMED
COMMENT Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.

FEATURES
source
1. .771
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="mynorm"
/note="Organ: Prostate; Vector: pBluescript; Directional cDNA library was constructed using Lambda ZP II kit (Stratagene). mRNA was extracted from a frozen normal prostate tissue (Mayo Clinic)."

ORIGIN
Query Match 16.3%; Score 328.2; DB 9; Length 771;
Best Local Similarity 70.9%; Pred. No. 1.9e-55;
Matches 497; Conservative 0; Mismatches 162; Indels 42; Gaps 5;

QY 781 ATGTTTTTTTAACTTACTATAAAGCCATGCAAAATGACAAAATCTATATTACTTTCTCA 840
|||
13 AGGTTTTTTTAACTTACTATAAAGCCATGCAAAATGACAAAATCTATATTACTTTCTCA 72
|||
841 AAATGGACCCCAAGAAACTTTGATTACTGTTCTTAACTGCCCTAATCTTAAATTACAGGA 900
|||
73 AAATGGACCCCAAGAAACTTTGATTACTGTTCTTAACTGCCCTAATCTTAAATTACAGGA 132
|||
901 ACTGTGCATCAGCTATTATGATTTCTATAAGCTATTTTCAAGCAGAAATGAGATAATAAATCC 960
|||
133 ACTGTGCATCAGCTATTATGATTTCTATAAGCTATTTTCAAGCAGAAATGAGATAATAAATCC 192
|||
961 AATGCTTTGATTGTTCTAGAAAGTATAGTAATTTGTTTCTTAAGGTGGTTCAGGCATCTA 1020
|||
193 AATGCTTTGATTGTTCTAGAAAGTATAGTAATTTGTTTCTTAAGGTGGTTCAGGCATCTA 252
|||
1021 CTCTTTTATCATTTACTTCAAAATGACATTTGCTAAAGACTGCATTTATTTTACTACTGTA 1080
|||
253 CTCTTTTATCATTTACTTCAAAATGACATTTGCTAAAGACTGCATTTATTTTACTACTGTA 312
|||

Db 445 GTGGGTGCTCTATTCCATCATCAGAACTTCTAGAACCCACTGGCGGATATGGTCTAAA 504
QY 570 ACGTGAGCTTGAGAGCTCTCTACTTAGGATGGACCAACGGCACTGGTGTGATTTGG 629
Db 505 GCGCCAGCTGGAGAGCCCTCTACATAAGCTGGACCAAAATCGCTGGGCTGATCGGTTG 564
QY 630 AGGAGCTCTGTTCTGCTGGGTTTTTTTGTGCAACGAAAGAGCAGTAGCTACAGATACTC 689
Db 565 AAGAGCACTGTTCTGAAGAGGGGTTGTGCAACTTGAAGGAGCAAAAAAATTCGTGTAATT 624
QY 690 GATACCTTCCCATCGCAC 707
Db 625 GGGACCATCCAATGGAC 642
RESULT 48
AY400666 325 bp DNA linear GSS 15-DEC-2003
LOCUS Pan troglodytes CLDN8 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY400666
VERSION AY400666.1 GI:39756655
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 325)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarimal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 325)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarimal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES Location/Qualifiers
source 1..325
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
gene <1..>325
/gene="CLDN8"
/locus_tag="HCM0633"
ORIGIN
Query Match 16.1%; Score 323.4; DB 29; Length 325;
Best Local Similarity 99.7%; Pred. No. 2.1e-54;
Matches 324; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 450 TCTGCTGACGGCTGGAATCATCTTTCATCATCACGGGCATGGTGGTCTCATCCCTGTGAG 509
Db 1 TCTGCTGACGGCTGGAATCATCTTTCATCATCACGGGCATGGTGGTCTCATCCCTGTGAG 60
QY 510 CTGGGTTGCCAATGCCATCATCAGAGATTTCTATAACTCAATAGTGAATGTTGCCAAAAA 569
Db 61 CTGGGTTGCCAATGCCATCATCAGAGATTTCTATAACCCCAATAGTGAATGTTGCCAAAA 120
QY 570 ACGTGAGCTTGAGAGAGCTCTCTACTTAGGATGGACCAACGGCACTGGTGTGATTTGG 629
Db 121 ACGTGAGCTTGAGAGAGCTCTCTACTTAGGATGGACCAACGGCACTGGTGTGATTTGG 180

QY 630 AGGAGCTCTGTTCTGCTGCGTTTTTTTGTGCAACGAAAGAGCAGTAGCTACAGATACTC 689
Db 181 AGGAGCTCTGTTCTGCTGCGTTTTTTTGTGCAACGAAAGAGCAGTAGCTACAGATACTC 240
QY 690 GATACCTTCCCATCGCACAAACCAAAAAGTTTATCACACCGGAAAGAACTCACCGAGCGT 749
Db 241 GATACCTTCCCATCGCACAAACCAAAAAGTTTATCACACCGGAAAGAACTCACCGAGCGT 300
QY 750 CTACTCCAGAGTCAGTATGTGTAG 774
Db 301 CTACTCCAGAGTCAGTATGTGTAG 325
RESULT 49
BF151294 578 bp mRNA linear EST 29-DEC-2000
LOCUS uz14e04.y1 NCI CGAP Mam5 Mus musculus cDNA clone IMAGE:3669054 5'
DEFINITION similar to SW:CLD8_MOUSE Q9Z260 CLAUDIN-8. ;, mRNA sequence.
ACCESSION BF151294
VERSION BF151294.1 GI:11032689
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 578)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml
MGI:1429822
Seq primer: -40RP from Gibco
High quality sequence stop: 409.
FEATURES Location/Qualifiers
source 1..578
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3669054"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/clone_lib="NCI CGAP Mam5"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"
ORIGIN
Query Match 15.9%; Score 319.8; DB 10; Length 578;
Best Local Similarity 81.0%; Pred. No. 9.7e-54;
Matches 372; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 90 GAGGATAATGGCAACCCATGCCTTAGAAATCGCTGGGCTGTTTCTTGGTGGTGAAT 149
Db 119 GAGGACGATGGCAACCTACGCTCTTCAAATGGCTGCACTGGTGGTGGTGGCAT 178
QY 150 GGTGGGCACAGTGGCTGTCACTGTCAATGCCTCAGTGGAGAGTGTGGGCCTTCAATGAAA 209
Db 179 GGTGGGCACGGTGGCTGTGACTATCATGCCTCAGTGGAGAGTGTCTGCCCTTCATCGAAG 238

QY	210	CAACATCGTGGTTTTTGAAAACTTCTGGGAAGGACTGTGGATGAATTGC GTGAGCAGGC	269
Db	239	TAACATTTGGTGTTTTGAGAACCCTGGGAAGGCTTGTGGATGAATTGTATGAGGCATGC	298
QY	270	TAACATCAGGATGCAGTGCAAAATCTATGATTCCTCTGCTGGCTCTTTCTCCGACCTACA	329
Db	299	CAACATCAAAATGCAGTGCAAGGTCTACGACTCCCTGCTGGCTCTTAGTCAGACCTCCA	358
QY	330	GGCAGCCAGAGGACTGATGTGTGCTGCTTCCGTGATGTCTTCTTGGCTTTTCATGATGC	389
Db	359	GGCATCCGAGGACTGATGTGTGCTGCTGCCCTCTTGGGTTTCTTGGCTTTTCATGACAGC	418
QY	390	CATCCTTGGCATGAAATGCACCAGGTGCACGGGGGACAATGAGAAGGTGAAGGCTCACAT	449
Db	419	CATCTCGGAATGGAGTGCAACAGATGCACCGGGGACAATGAGAACGTGAAGAGCCGCAT	478
QY	450	TCTGCTGACGGCTGGAATCATCTTTCATCATCACGGGCATGGTGGTCTCATCCCTGTGAG	509
Db	479	TCTTGCTACAGCCC GAATCATCTTTCTTAACACCGGCTTTGTGTGCTCATCCCTGTGAG	538
QY	510	CTGGGTGCCCAATGCCATCATCAGAGATTTCATAA CTC	548
Db	539	CTTGACTGCCAATTC CATCATCAGAGACTTCTACACCC	577

```

RESULT 50
AA769736/c
LOCUS
DEFINITION
AA769736
348 bp . mRNA linear EST 08-FEB-1998
ob20c04.s1 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1324230 3',
mRNA sequence.
AA769736
AA769736.1 GI:2820974
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 348)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
unknown library type
Insert Length: 1460 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 223.
Location/Qualifiers
1. .348
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1324230"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/clone_lib="NCI CGAP Kid5"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
AACTGGAAGAAATCGCGCGCAATATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo. "
FEATURES
source

```

Query Match 15.5%; Score 311.2; DB 9; Length 348;
Best Local Similarity 99.1%; Pred. NO. 5.7e-52;
Matches 313; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1031 CAATTACTTCAAAATGACATTGCTAAAGACTGCATTATTTTACTACTGTAAATTTCTCCAC 1090

Db	348	CATTTACTTCAAAATGACATTGCTAAAGACTGCATTATTTTACTACTGTAATTTCTCCAC	289
QY	1091	GACATAGCATTATGTACATAGATGAGTGTAAACATTTATATCTCACATAGAGACATGCTTA	1150
Db	288	GACATAGCATTATGTACATAGATGAGTGTAAACATTTATATCTCACATAGAGACATGCTTA	229
QY	1151	TATGGTTTTATTTAAATGAAATGCCAGTCCATTACACTGAATAAATAGAACTCAACTAT	1210
Db	228	TATGGTTTTATTTAAATGAAATGCCAGTCCATTACACTGAATAAATAGAACTCAACTAT	169
QY	1211	TGCTTTTTCAGGGAAATCATGGATAGGGTTGAAGAAGTTACTATTAAATGTTTAAAAACA	1270
Db	168	TGCTTTTTCAGGGAAATCATGGATAGGGTTGAAGAAGTTACTATTAAATGTTTAAAAACA	109
QY	1271	GCTTAGGGATTAAATGTCCTCCATTTATAATGAAGATTAAAAATGAAGGCTTTTAATCAGCAT	1330
Db	108	GCTTAGGGATTAAATGTCCTCCATTTATAATGAAGATTAAAAATGAAGGCTTTTAATCAGCAT	49
QY	1331	TGTAAGGAAATTGAA	1346
Db	48	TGTAAGGAAACATGCA	33

Search completed: September 3, 2004, 00:56:25
Job time : 3563 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 3, 2004, 01:55:56 ; Search time 2675 Seconds
(without alignments)
3645.676 Million cell updates/sec

Title: US-10-063-732-120
Perfect score: 1172
Sequence: 1 MATHALEIAGLFLGGVMVG.....QKSYHTGKKSPSVYRSQYV 225

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2 1/USPTO.spool/US10063732/runat_01092004_155516_20440/app_query.fasta_1.391
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10063732 @CGN 1 1 2527 @runat_01092004_155516_20440 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1172	100.0	1700	9	BC058004 Homo sapi
2	1172	100.0	1835	9	BC020866 Homo sapi
3	1172	100.0	1931	9	AJ250711 Homo sapi
4	1172	100.0	2010	6	AX092388 Sequence
5	1172	100.0	2010	6	AX376290 Sequence
6	1172	100.0	2010	6	AX697259 Sequence
7	1172	100.0	2010	9	AY358707 Homo sapi
8	1172	100.0	81583	2	AC131927 Homo sapi
9	1172	100.0	137955	9	AP000884 Homo sapi
10	1172	100.0	340000	9	AP001707 Homo sapi
11	1167	99.6	176580	9	BS000178 Pan trogl
12	1167	99.6	267172	9	BS000177 Pan trogl
13	1140	97.3	837	6	AX866683 Sequence
14	1140	97.3	837	6	BD146745 Primer fo
15	1140	96.8	191923	2	AP001846 Homo sapi
16	1127	96.2	1890	6	AX879004 Sequence
17	1127	96.2	1890	6	BD157589 Primer fo
18	1127	96.2	1890	9	AK022269 Homo sapi
19	1106	94.4	1002	6	AX497202 Sequence
20	1018	86.9	240479	2	AC096377 Rattus no
21	1012	86.3	678	6	E31609 Tight junct
22	1012	86.3	678	10	AF087826 Mus muscu
23	1012	86.3	2325	10	BC003868 Mus muscu
24	1012	86.3	217278	2	AC113190 Mus muscu
25	1012	86.3	256720	10	AC110241 Mus muscu
26	696	59.4	675	9	HSA250712 Homo sapi
27	696	59.4	833	6	AX497204 Sequence
28	676	57.7	1160	6	AX697263 Sequence
29	676	57.7	1160	9	AY358094 Homo sapi
30	663	56.6	615	6	AR414198 Sequence
31	663	56.6	615	6	BD109751 EST and e
32	542.5	46.3	902	5	BC059750 Silurana
33	538	45.9	916	5	AF359435 Xenopus l
34	538	45.9	925	5	AB072909 Xenopus l
35	538	45.9	929	5	AB072908 Xenopus l
36	538	45.9	1736	5	AF334677 Gallus ga
37	534	45.6	660	6	BD076388 Human.pro
38	534	45.6	1250	6	AX302529 Sequence
39	534	45.6	1250	9	AB000714 Homo sapi
40	534	45.6	1279	6	BD076398 Human.pro
41	534	45.6	1294	9	BC016056 Homo sapi
42	534	45.6	1601	9	AF007189 Homo sapi
43	534	45.6	148689	9	AC093168 Homo sapi
44	533	45.5	895	5	AF224712 Xenopus l
45	528	45.1	857	5	XLAJ9313 Xenopus l

ALIGNMENTS

RESULT 1

BC058004	1700 bp	mRNA	linear	PRI 07-OCT-2003
LOCUS	Homo sapiens claudin 8, mRNA (cDNA clone MGC:61830 IMAGE:4606880), complete cds.			
DEFINITION	Homo sapiens (human)			
ACCESSION	BC058004.1 GI:34783654			
VERSION	MGC.			
KEYWORDS	Homo sapiens (human)			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 1700)			
	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.			
	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences			
TITLE	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)			
JOURNAL	22388257			
MEDLINE	12477932			
PUBMED	2 (bases 1 to 1700)			
REFERENCE	Strausberg,R.			
AUTHORS	Direct Submission			
TITLE	Submitted (08-SEP-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
JOURNAL	NIH-MGC Project URL: http://mgc.nci.nih.gov			
REMARK	Contact: MGC help desk			
COMMENT	Email: cgapbs-r@mail.nih.gov			
	Tissue Procurement: CLONTECH			
	cDNA Library Preparation: CLONTECH Laboratories, Inc.			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
	DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),			
	Gaithersburg, Maryland;			
	Web site: http://www.nisc.nih.gov/			
	Contact: nisc_mgc@nhgri.nih.gov			
	Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.			
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov			
	Series: IRAL Plate: 48 Row: d Column: 13			
	This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21314655.			
FEATURES	Location/Qualifiers			
source	1..1700			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="MGC:61830 IMAGE:4606880"			
	/tissue_type="Kidney"			
	/clone_lib="NIH_MGC_75"			
	/lab_host="DH10B"			
	/note="Vector: pDNR-LIB"			
gene	1..1700			
	/gene="CLDN8"			
	/db_xref="LocusID:9073"			
CDS	76..753			
	/codon_start=1			
	/product="CLDN8 protein"			
	/protein_id="AAH58004.1"			
	/db_xref="GI:34783655"			
	/db_xref="LocusID:9073"			
	/translation="MATHALEIAGLFLGGVGMVGTAVTVMPOQRVSAPIENNVVFE NFEGLWMNCVRQANIRMOCKIYDSLALSPLQAARGLMCAASVMSFLAFMAILGM KCTRCTGDNEKVKAHILLTAGIIFIITGMVVLIPVSWVANAIIRDFYNSIVNVAQKRE LGEALYLGTWTTALVLIVGALFCCVFCCKNEKSSSYRYSIPSHRTTQKSYHTGKSPSV YSRSQYV"			
misc_feature	160..621			
	/note="PMP22_Claudin; Region: PMP-22/EMP/MP20/Claudin family"			
	/db_xref="CDD:pfam00822"			
ORIGIN				
Alignment Scores:				
Pred. No.:	5,53e-123	Length:	1700	
Score:	1172.00	Matches:	225	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	100.00%	Indels:	0	
DB:	9	Gaps:	0	
US-10-063-732-120 (1-225) x BC058004 (1-1700)				
Qy	1	MetAlaThrHisAlaLeuGluIleAlaGlyLeuPheLeuGlyValGlyMetValGly	20	
Db	76	ATGGCAACCCATGCCTTAGAAATCGCTGGGCTGTTCTTGGTGGTGGTGGC	135	
Qy	21	ThrValAlaValThrValMetProGlnTrpArgValSerAlaPheIleGluAsnAsnIle	40	
Db	136	ACAGTGGTGTCACTGTCACTGCCTCATGCCCTCAGTGGAGAGTGTGGCCTTCATTGAAACCAACATC	195	
Qy	41	ValValPheGluAsnPheTrpGluGlyLeuTrpMetAsnCysValArgGlnAlaAsnIle	60	
Db	196	GTGGTTTTGGAAACTTCTGGGAAGGACTGTGGATGAATGCTGAGGCAAGCTAACATC	255	
Qy	61	ArgMetGlnCysLysIleTyrAspSerLeuLeuAlaLeuSerProAspLeuGlnAlaAla	80	
Db	256	AGGATGCAGTGCAAAATCTATGATCCCTGCTGGCTCTTCTCCGGACCTACAGGCAGCC	315	
Qy	81	ArgGlyLeuMetCysAlaAlaSerValMetSerPheLeuAlaPheMetMetAlaIleLeu	100	
Db	316	AGAGGACTGATGTGTGCTGCTTCCGTGATGTCTCTTCTGGCTTTCATGATGGCCATCCTT	375	
Qy	101	GlyMetLysCysThrArgCysThrGlyAspAsnGluLysValLysAlaHisIleLeuLeu	120	
Db	376	GGCATGAAATGCACAGGTGCACGGGGGACAAATGAGAGGTGAAGGCTCACATTCGCTG	435	
Qy	121	ThrAlaGlyIleIlePheIleIleThrGlyMetValValLeuIleProValSerTrpVal	140	
Db	436	ACGGCTGGAATCATCTTCATCATCACGGGCATGGTGGTGTCTCATCCCTGTGAGCTGGTT	495	
Qy	141	AlaAsnAlaIleIleArgAspPheTyrAsnSerIleValAsnValAlaGlnLysArgGlu	160	
Db	496	GCCAATGCCATCATCAGAGATTTCTATACTCACTAGTAGAATGTTGCCCAAAACGTGAG	555	
Qy	161	LeuGlyGluAlaLeuTyrLeuGlyTrpThrThrAlaLeuValLeuIleValGlyGlyAla	180	
Db	556	CTTGGAGAGAGCTCTCTACTTAGGATGGACCAACGGCACTGGTGTGCTGATTGTGGAGGAGCT	615	
Qy	181	LeuPheCysCysValPheCysCysAsnGluLysSerSerTyrArgTyrSerIlePro	200	
Db	616	CTGTTCTGCTGCGTTTTTTTGTGCAACGAAAGACAGTAGCTACAGATACTCGATACCT	675	

QY 201 SerHisArgThrThrGlnLysSerTyrHisThrGlyLysLysSerProSerValTyrSer 220
|||||
Db 676 TCCCATCGCACAAACCCAAAAAGTTATCACACGGAAAGAGTCACCGAGCGTCTACTCC 735

QY 221 ArgSerGlnTyrVal 225
|||||
Db 736 AGAAGTCAGTATGTG 750

RESULT 2
BC020866 1835 bp mRNA linear PRI 06-OCT-2003
LOCUS Homo sapiens claudin 8, mRNA (cdna clone MGC:24067 IMAGE:4594155),
DEFINITION complete cds.
ACCESSION BC020866
VERSION BC020866.1 GI:18089189
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1835)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 1835)
Strausberg,R.
Direct Submission
Submitted (03-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 36 Row: e Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21314655.
Location/Qualifiers
1. .1835
/organism="Homo sapiens"

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:24067 IMAGE:4594155"
/tissue_type="Kidney"
/clone_lib="NIH_MGC_75"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
1. .1835
/gene="CLDN8"
/db_xref="LocusID:9073"
110. .787
/codon_start=1
/product="claudin 8"
/protein_id="AAH20866.1"
/db_xref="GI:18089190"
/db_xref="LocusID:9073"
/translation="MATHALEIAGFLGGVGMVGTVAVTMPQWRVSAFIENNIVVEE
NFWEGLMNCVRQANIRMQCKIYDSLALSPDLQAARGLMCAASVMSFLAFMMAILGM
KCTRCTGDNEKVKAHILLTAGIIFIITGMVVLIPVSWVANAIIPDFNSIVNVAQKRE
LGEALYLGWTTALVLIVGALFCVFCCKSEKSSSYRYSIPSHRTTQKSYHTGKSPSV
YRSRQIV"
194. .655
/note="PMP22_Claudin; Region: PMP-22/EMP/MP20/Claudin
family"
/db_xref="CDD:pfam00822"

ORIGIN
Alignment Scores:
Pred. No.: 6.09e-123 Length: 1835
Score: 1172.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-10-063-732-120 (1-225) x BC020866 (1-1835)

QY 1 MetAlaThrHisAlaLeuGluIleAlaGlyLeuPheLeuGlyGlyValGlyMetValGly 20
|||||
Db 110 ATGGCAACCCATGCCCTTAGAAATCGCTGGCGTGTTCCTGTGGTGTGGAATGGTGGC 169

QY 21 ThrValAlaValThrValMetProGlnTrpArgValSerAlaPheIleGluAsnIle 40
|||||
Db 170 ACAGTGGCTGTCACTGTCTATGCCCTCAGTGGAGAGTGTGGCGCTTCATTGAAAAACAACATC 229

QY 41 ValValPheGluAsnPheTrpGluGlyLeuTrpMetAsnCysValArgGlnAlaAsnIle 60
|||||
Db 230 GTGGTTTTTGAATACTTCTGGAGAGACTGTGGATGAATTCGTGAGGCAGGCTAACATC 289

QY 61 ArgMetGlnCysLysIleTyrAspSerLeuLeuAlaLeuSerProAspLeuGlnAlaAla 80
|||||
Db 290 AGGATGCAGTGCAAAATCTATGATTCCTCTGCTGGCTCTTCTCCGACCTACAGGCAGCC 349

QY 81 ArgGlyLeuMetCysAlaAlaSerValMetSerPheLeuAlaPheMetMetAlaIleLeu 100
|||||
Db 350 AGAGGACTGATGTGTGCTGCTTCCTGATGTCTCTTCTGGCTTTCATGATGGCCATCCTT 409

QY 101 GlyMetLysCysThrArgCysThrGlyAspAsnGluLysValLysAlaHisIleLeuLeu 120
|||||
Db 410 GGCATGAATGCACAGGTGCACGGGGACAAATGAGAAGGTGAAGGCTCACATTTCTGCTG 469

QY 121 ThrAlaGlyIleIlePheIleIleThrGlyMetValValLeuIleProValSerTrpVal 140
|||||
Db 470 ACGGCTGGAATCATCTTCATCATCATCAGGGCATGGTGGTGCTCATCCCTGTGAGCTGGTT 529

QY 141 AlaAsnAlaIleIleArgAspPheTyrAsnSerIleValAsnValAlaGlnLysArgGlu 160
|||||
Db 530 GCCAATGCCATCATCAGAGATTTCTATAACTCAATAGTAGTATGTTGCCCCAAAACGTGAG 589

QY 161 LeuGlyGluAlaLeuTyrLeuGlyTrpThrThrAlaLeuValLeuIleValGlyGlyAla 180
|||||
Db 590 CTTGGAGAAGCTCTCTACTTAGGATGGACCAACGGCACTGGTGTGCTGATTGTTGGAGGAGCT 649

Query Match: 100.00% Indels: 0 DB: 6 Gaps: 0

US-10-063-732-120 (1-225) x AX092388 (1-2010)

QY 1 MetAlaThrHisAlaLeuGluIleAlaGlyLeuPheLeuGlyValGlyMetValGly 20
Db 97 ATGGCAACCCATGCCCTTAGAAATCGCTGGCTGTTCTTGGTGGTGTGGAAATGGTGGGC 156

QY 21 ThrValAlaValThrValMetProGlnTrpArgValSerAlaPheIleGluAsnIle 40
Db 157 ACAGTGGCTGTCACTGTCTATGCTCAGTGGAGAGTGTGGCCTTCATTGAAAAACAACATC 216

QY 41 ValValPheGluAsnPheTrpGluGlyLeuTrpMetAsnCysValArgGlnAlaAsnIle 60
Db 217 GTGGTTTTGAAAACTCTGGGAAGGACTCTGGATGAATGGGTGAGGCAGGCTAACATC 276

QY 61 ArgMetGlnCysLysIleTyrAspSerLeuLeuAlaLeuSerProAspLeuGlnAlaAa 80
Db 277 AGGATGCAGTGCAAAATCTATGATTCCTCGTGGCTCTTCTCCGGACCTACAGGCAGCC 336

QY 81 ArgGlyLeuMetCysAlaAlaSerValMetSerPheLeuAlaPheMetMetAlaIleLeu 100
Db 337 AGAGGACTGATGTGTGCTTCCGTGATCTCTTGGCTTCATGATGGCCATCCTT 396

QY 101 GlyMetLysCysThrArgCysThrGlyAspAsnGluLysValIysAlaHisIleLeuLeu 120
Db 397 GGCATGAATGCACCAAGTGCACGGGGACAATGAGAAGGTGAAGGCTCACATCTGCTG 456

QY 121 ThrAlaGlyIleIlePheIleIleThrGlyMetValValIleProValSerTrpVal 140
Db 457 ACGCTGGAATCATCTTCATCATCACGGGCATGGTGTCTCATCCCTGTGAGCTGGGTT 516

QY 141 AlaAsnAlaIleIleArgAspPheTyrAsnSerIleValAsnValAlaGlnLysArgGlu 160
Db 517 GCCAATGCCATCATCAGAGATTCTATACTCAATAGTGAATGTTGCCCAAAACGTGAG 576

QY 161 LeuGlyGluAlaLeuTyrLeuGlyTrpThrAlaLeuValIleValGlyGlyAla 180
Db 577 CTTGAGAAAGCTCTCTACTTAGGATGGACCGGCACCTGGTGTGATTTGTTGGAGAGCT 636

QY 181 LeuPheCysCysValPheCysAsnGluLysSerSerSerTyrArgTyrSerIlePro 200
Db 637 CTGTTCTGCTGCTTTTGTGACGAAAGAGCAGTAGCTACAGATACTCGATACCT 696

QY 201 SerHisArgThrThrGlnLysSerTyrHisThrGlyLysSerProSerValTyrSer 220
Db 697 TCCCATCGCACAAACCCAAAAAGTTATCACACCGGAAAGAGTACCGGCGTCTACTCC 756

QY 221 ArgSerGlnTyrVal 225
Db 757 AGAAGTCAGTATGTG 771

RESULT 5
AX376290
LOCUS AX376290 2010 bp DNA linear PAT 01-MAR-2002
DEFINITION Sequence 357 from Patent WO0168848.
ACCESSION AX376290
VERSION AX376290.1 GI:19170536
KEYWORDS
SOURCE Homo sapiens (human).
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS Baker,K.P., Chen,J., Desnoyers,L., Goddard,A., Godowski,P.J.,
Gurney,A.L., Pan,J., Smith,V., Watanabe,C.K., Wood,W.I. and
Zhang,Z.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0168848-A 357 20-SEP-2001;
Genentech, Inc. (US)
FEATURES Location/Qualifiers

source 1. .2010
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores: 6.83e-123 Length: 2010
Score: 1172.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-063-732-120 (1-225) x AX376290 (1-2010)

QY 1 MetAlaThrHisAlaLeuGluIleAlaGlyLeuPheLeuGlyValGlyMetValGly 20
Db 97 ATGGCAACCCATGCCCTTAGAAATCGCTGGCTGTTCTTGGTGGTGTGGAAATGGTGGGC 156

QY 21 ThrValAlaValThrValMetProGlnTrpArgValSerAlaPheIleGluAsnIle 40
Db 157 ACAGTGGCTGTCACTGTCTATGCTCAGTGGAGAGTGTGGCCTTCATTGAAAAACAACATC 216

QY 41 ValValPheGluAsnPheTrpGluGlyLeuTrpMetAsnCysValArgGlnAlaAsnIle 60
Db 217 GTGGTTTTGAAAACTCTGGGAAGGACTGTGGATGAATGGGTGAGGCAGGCTAACATC 276

QY 61 ArgMetGlnCysLysIleTyrAspSerLeuLeuAlaLeuSerProAspLeuGlnAlaAa 80
Db 277 AGGATGCAGTGCAAAATCTATGATTCCTCGTGGCTCTTCTCCGGACCTACAGGCAGCC 336

QY 81 ArgGlyLeuMetCysAlaAlaSerValMetSerPheLeuAlaPheMetMetAlaIleLeu 100
Db 337 AGAGGACTGATGTGTGCTTCCGTGATGTCTTGGCTTTCATGATGGCCATCCTT 396

QY 101 GlyMetLysCysThrArgCysThrGlyAspAsnGluLysValIysAlaHisIleLeuLeu 120
Db 397 GGCATGAATGCACCAAGTGCACGGGGACAATGAGAAGGTGAAGGCTCACATCTGCTG 456

QY 121 ThrAlaGlyIleIlePheIleIleThrGlyMetValValIleProValSerTrpVal 140
Db 457 ACGCTGGAATCATCTTCATCATCACGGGCATGGTGTGCTCATCCCTGTGAGCTGGGTT 516

QY 141 AlaAsnAlaIleIleArgAspPheTyrAsnSerIleValAsnValAlaGlnLysArgGlu 160
Db 517 GCCAATGCCATCATCAGAGATTCTATACTCAATAGTGAATGTTGCCCAAAACGTGAG 576

QY 161 LeuGlyGluAlaLeuTyrLeuGlyTrpThrAlaLeuValIleValGlyGlyAla 180
Db 577 CTTGAGAAAGCTCTCTACTTAGGATGGACCGGCACCTGGTGTGATTTGTTGGAGAGCT 636

QY 181 LeuPheCysCysValPheCysAsnGluLysSerSerSerTyrArgTyrSerIlePro 200
Db 637 CTGTTCTGCTGCTTTTGTGCAACGAAAGAGCAGTAGCTACAGATACTCGATACCT 696

QY 201 SerHisArgThrThrGlnLysSerTyrHisThrGlyLysSerProSerValTyrSer 220
Db 697 TCCCATCGCACAAACCCAAAAAGTTATCACACCGGAAAGAGTACCGGCGTCTACTCC 756

QY 221 ArgSerGlnTyrVal 225
Db 757 AGAAGTCAGTATGTG 771

RESULT 6
AX697259
LOCUS AX697259 2010 bp DNA linear PAT 02-APR-2003
DEFINITION Sequence 327 from Patent WO0078961.
ACCESSION AX697259
VERSION AX697259.1 GI:29498418
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE	Db	757	AGAAGTCAGTATGTG	771
	RESULT 7			
	AY358707			
	LOCUS	AY358707	2010 bp	mRNA linear PRI 03-OCT-2003
	DEFINITION	Homo sapiens clone DNA73735 CLDN8 (UNQ779) mRNA, complete cds.		
	ACCESSION	AY358707		
	VERSION	AY358707.1	GI:37182535	
	KEYWORDS	FLI_CDNA.		
	SOURCE	Homo sapiens (human)		
	ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
	1	(bases 1 to 2010)		
	AUTHORS	Clark,H.F., Gurney,A.L., Abaya,E., Baker,K., Baldwin,D., Brush,J., Chen,J., Chow,B., Chui,C., Crowley,C., Currell,B., Deuel,B., Dowd,P., Eaton,D., Foster,J., Grimaldi,C., Gu,Q., Hass,P.E., Heldens,S., Huang,A., Kim,H.S., Klimowski,L., Jin,Y., Johnson,S., Lee,J., Lewis,L., Liao,D., Mark,M., Robbie,E., Sanchez,C., Schoenfeld,J., Seshagiri,S., Simmons,L., Singh,J., Smith,V., Stinson,J., Vagts,A., Vandlen,R., Watanabe,C., Wieand,D., Woods,K., Xie,M.H., Yansura,D., Yi,S., Yu,G., Yuan,J., Zhang,M., Zhang,Z., Goddard,A., Wood,W.I. and Godowski,P.		
	TITLE	The Secreted Protein Discovery Initiative (SPDI), a Large-Scale Effort to Identify Novel Human Secreted and Transmembrane Proteins: A Bioinformatics Assessment		
	JOURNAL	Genome Res. 13 (10), 2265-2270 (2003)		
	PUBMED	12975309		
		2 (bases 1 to 2010)		
REFERENCE	Clark,H.F.			
	Direct Submission			
TITLE	Submitted (01-AUG-2003) Department of Bioinformatics, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA			
	Location/Qualifiers			
FEATURES	1. .2010			
	/organism="Homo sapiens"			
source	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
gene	/clone="DNA73735"			
	1. .2010			
CDS	/locus_tag="UNQ779"			
	97. .774			
CDS	/locus tag="UNQ779"			
	/note="PRO1573"			
CDS	/codon_start=1			
	/product="CLDN8"			
CDS	/protein_id="AAQ89070.1"			
	/db_xref="GI:37182536"			
CDS	/translation="MATHALEIAGLFLGGVMGTVAVTVMQWRVSAFIENNIIVVFE			
	NFEGLMNCVQRQANIRMQCKIYDLSLLSPDLQAARGLMCAASVMSFLAFMAILGM			
CDS	KCTRCTGNEKVKAHILLTAGIIFIITGMVLIIPVSWANAIIRDYFNIVNVAQKRE			
	LGEALYLGWTTLALVIVGALFCCVFCCKNEKSSSYRISIPSHRTTQKSYHTGKXSPSV			
CDS	YRSQYV"			
ORIGIN	Alignment Scores:			
	Pred. No.: 6.83e-123 Length: 2010			
ORIGIN	Score: 1172.00 Matches: 225			
	Percent Similarity: 100.00% Conservative: 0			
ORIGIN	Best Local Similarity: 100.00% Mismatches: 0			
	Query Match: 100.00% Indels: 0			
ORIGIN	DB: 6 Gaps: 0			
	US-10-063-732-120 (1-225) x AX697259 (1-2010)			
ORIGIN	Qy 1 MetAlaThrHisAlaLeuGluIleAlaGlyLeuPheLeuGlyGlyValGlyMetValGly 20			
	Db 97 ATGGCAACCCATGCCTTAGAAATCGCTGGCTGTTCTTGGTGGTGGATGGTGGC 156			
ORIGIN	Qy 21 ThrValAlaValThrValMetProGlnTrpArgValSerAlaPheIleGluAsnAsnIle 40			
	Db 157 ACAGTGGCTGTCACTGTCTATGCCCTCAGTGGAGAGTGTGGCTTCAITGAAACAACATC 216			
ORIGIN	Qy 41 ValValPheGluAsnPheTrpGluGlyLeuTrpMetAsnCysValArgGlnAlaAsnIle 60			
	Db 217 GTGGTTTTTGAAAACTTCTGGGAAGGACTGTGGATGAATTCGTGAGCAGGCTAACATC 276			
ORIGIN	Qy 61 ArgMetGlnCysLysIleTyrAspSerLeuLeuAlaLeuSerProAspLeuGlnAlaAla 80			
	Db 277 AGGATGCAGTGCAGAAATCTATGATTCCTGTGGTCTTCTCCGGACCTACAGGCAGCC 336			
ORIGIN	Qy 81 ArgGlyLeuMetCysAlaAlaSerValMetSerPheLeuAlaPheMetMetAlaIleLeu 100			
	Db 337 AGAGGACTGATGTGTCTTCCGTGATGTCTTCTTGGCTTTTCATGATGGCCATCCTT 396			
ORIGIN	Qy 101 GlyMetLysCysThrArgCysThrGlyAspAsnGluLysValLysAlaHisIleLeuLeu 120			
	Db 397 GGCATGAATGCACAGGTGCACGGGGACAATGAGAAGGTGAAGGCTCACATTCGTCTG 456			
ORIGIN	Qy 121 ThrAlaGlyIleIlePheIleIleThrGlyMetValValLeuIleProValSerTrpVal 140			
	Db 457 ACGGCTGGAATCATCTTCATCATCACGGGCATGGTGGTGTCTATCCCTGTGAGCTGGGT 516			
ORIGIN	Qy 141 AlaAsnAlaIleIleArgAspPheTyrAsnSerIleValAsnValAlaGlnLysArgGlu 160			
	Db 517 GCCAATGCCATCATCAGAGATTTCTATACTCAATAGTGAATGTTGCCCAAAACGTGAG 576			
ORIGIN	Qy 161 LeuGlyGluAlaLeuTyrLeuGlyTrpThrThrAlaLeuValLeuIleValGlyGlyAla 180			
	Db 577 CTTGGAGAAGCTCTCTACTTAGGATGGACCAACGGCACTGGTGTGATTGTTGGAGGCT 636			
ORIGIN	Qy 181 LeuPheCysCysValPheCysCysAsnGluLysSerSerTyrArgTyrSerIlePro 200			
	Db 637 CTGTTCTGCTGCGTTTTTTGTTGCCAACGAAAGAGACAGTAGCTACAGATACTCGATACCT 696			
ORIGIN	Qy 201 SerHisArgThrThrGlnLysSerTyrHisThrGlyLysLysSerProSerValTyrSer 220			
	Db 697 TCCCATCGCACAAACCCAAAAAGTTATCACACCGGAAGAAGTCAACCGAGCGTCTACTCC 756			
ORIGIN	Qy 221 ArgSerGlnTyrVal 225			

QY 41 ValValPheGluAsnPheTrpGluGlyLeuTrpMetAsnCysValArgGlnAlaAsnIle 60
Db 217 GTGGTTTTGAAACTTCTGGGAAGGACTGTGGATGAATTGCGTGAGGCAGCTACATC 276

QY 61 ArgMetGlnCysLysIleTyrAspSerLeuLeuAlaLeuSerProAspLeuGlnAlaAla 80
Db 277 AGGATGAGTGCAGAAATCTATGATTCCTCGTGGCTCTTTCTCCGAGCCTACAGGCAGCC 336

QY 81 ArgGlyLeuMetCysAlaAlaSerValMetSerPheLeuAlaPheMetMetAlaIleLeu 100
Db 337 AGAGGACTGATGTGTGCTGCTCCGTGATGCTCTTCTTGGCTTTCATGATGGCCATCCTT 396

QY 101 GlyMetLysCysThrArgCysThrGlyAspAsnGluLysValLysAlaHisIleLeuLeu 120
Db 397 GGCATGAAATGCACAGGTGCACGGGGACATGAGAAGGTGAAGGCTCACATTCGTCTG 456

QY 121 ThrAlaGlyIleIlePheIleIleThrGlyMetValValLeuIleProValSerTrpVal 140
Db 457 ACGGCTGGAATCATCTTCATCATCATCAGGGCATGGTGGTGCTCATCCCTGTGAGCTGGGTT 516

QY 141 AlaAsnAlaIleIleArgAspPheTyrAsnSerIleValAsnValAlaGlnLysArgGlu 160
Db 517 GCCAATGCCATCATCAGAGATTTCTATACTCAATAGTAGTGAATGTTGCCCAAAACGTGAG 576

QY 161 LeuGlyGluAlaLeuTyrLeuGlyTyrThrAlaLeuValLeuIleValGlyGlyAla 180
Db 577 CTTGGAGAAGCTCTCTACTAGGATGGACCGCACTGGTGTGCTGATTTGGAGGAGCT 636

QY 181 LeuPheCysCysValPheCysCysAsnGluLysSerSerSerTyrArgTyrSerIlePro 200
Db 637 CTGTTCTGCTGCGTTTTTTTGTGTGCAACGAAAGAGCAGTAGCTACAGATACTCGATACCT 696

QY 201 SerHisArgThrThrGlnLysSerTyrHisThrGlyLysLysSerProSerValTyrSer 220
Db 697 TCCCATCGCACAAACCCAAAAGTTATCACACGGAAAGAGTCAACCGCGTCTACTCC 756

QY 221 ArgSerGlnTyrVal 225
Db 757 AGAAGTCAGTATGTG 771

RESULT 8
AC131927/c
LOCUS AC131927 81583 bp DNA linear HTG 30-DEC-2002
DEFINITION Homo sapiens chromosome 8 clone RP11-945D10 map 8, LOW-PASS
SEQUENCE SAMPLING.

ACCESSION AC131927
VERSION AC131927.2 GI:27414257
KEYWORDS HTG; HTGS PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 81583)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 8, clone RP11-945D10
Unpublished

REFERENCE 2 (bases 1 to 81583)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,

Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 81583)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-DEC-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 30, 2002 this sequence version replaced gi:22507216.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L28078
Center clone name: 945_D_10

* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 962: contig of 962 bp in length
* 963 1062: gap of 100 bp
* 1063 1998: contig of 936 bp in length
* 1999 2098: gap of 100 bp
* 2099 3025: contig of 927 bp in length
* 3026 3125: gap of 100 bp
* 3126 4085: contig of 960 bp in length
* 4086 4185: gap of 100 bp
* 4186 5143: contig of 958 bp in length
* 5144 5243: gap of 100 bp
* 5244 6191: contig of 948 bp in length
* 6192 6291: gap of 100 bp
* 6292 7252: contig of 961 bp in length
* 7253 7352: gap of 100 bp
* 7353 8293: contig of 941 bp in length
* 8294 8393: gap of 100 bp
* 8394 9328: contig of 935 bp in length
* 9329 9428: gap of 100 bp
* 9429 10373: contig of 945 bp in length
* 10374 10473: gap of 100 bp
* 10474 11394: contig of 921 bp in length

TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

* 11395 11494: gap of 100 bp
* 11495 12397: contig of 903 bp in length
* 12398 12497: gap of 100 bp
* 12498 13386: contig of 889 bp in length
* 13387 13486: gap of 100 bp
* 13487 14417: contig of 931 bp in length
* 14418 14517: gap of 100 bp
* 14518 15479: contig of 962 bp in length
* 15480 15579: gap of 100 bp
* 15580 16546: contig of 967 bp in length
* 16547 16646: gap of 100 bp
* 16647 17599: contig of 953 bp in length
* 17600 17699: gap of 100 bp
* 17700 18667: contig of 968 bp in length
* 18668 18767: gap of 100 bp
* 18768 19733: contig of 966 bp in length
* 19734 19833: gap of 100 bp
* 19834 20785: contig of 952 bp in length
* 20786 20885: gap of 100 bp
* 20886 21824: contig of 939 bp in length
* 21825 21924: gap of 100 bp
* 21925 22867: contig of 943 bp in length
* 22868 22967: gap of 100 bp
* 22968 23883: contig of 916 bp in length
* 23884 23983: gap of 100 bp
* 23984 24903: contig of 920 bp in length
* 24904 25003: gap of 100 bp
* 25004 25971: contig of 968 bp in length
* 25972 26071: gap of 100 bp
* 26072 27042: contig of 971 bp in length
* 27043 27142: gap of 100 bp
* 27143 28102: contig of 960 bp in length
* 28103 28202: gap of 100 bp
* 28203 29173: contig of 971 bp in length
* 29174 29273: gap of 100 bp
* 29274 30241: contig of 968 bp in length
* 30242 30341: gap of 100 bp
* 30342 31301: contig of 960 bp in length
* 31302 31401: gap of 100 bp
* 31402 32340: contig of 939 bp in length
* 32341 32440: gap of 100 bp
* 32441 33375: contig of 935 bp in length
* 33376 33475: gap of 100 bp
* 33476 34422: contig of 947 bp in length
* 34423 34522: gap of 100 bp
* 34523 35449: contig of 927 bp in length
* 35450 35549: gap of 100 bp
* 35550 36516: contig of 967 bp in length
* 36517 36616: gap of 100 bp
* 36617 37574: contig of 958 bp in length
* 37575 37674: gap of 100 bp
* 37675 38660: contig of 986 bp in length
* 38661 38760: gap of 100 bp
* 38761 39733: contig of 973 bp in length
* 39734 39833: gap of 100 bp
* 39834 40803: contig of 970 bp in length
* 40804 40903: gap of 100 bp
* 40904 41870: contig of 967 bp in length
* 41871 41970: gap of 100 bp
* 41971 42938: contig of 968 bp in length
* 42939 43038: gap of 100 bp
* 43039 43967: contig of 929 bp in length
* 43968 44067: gap of 100 bp
* 44068 45006: contig of 939 bp in length
* 45007 45106: gap of 100 bp
* 45107 46023: contig of 917 bp in length
* 46024 46123: gap of 100 bp
* 46124 47026: contig of 903 bp in length
* 47027 47126: gap of 100 bp
* 47127 48085: contig of 959 bp in length
* 48086 48185: gap of 100 bp
* 48186 49156: contig of 971 bp in length
* 49157 49256: gap of 100 bp

* 49257 50228: contig of 972 bp in length
* 50229 50328: gap of 100 bp
* 50329 51306: contig of 978 bp in length
* 51307 51406: gap of 100 bp
* 51407 52385: contig of 979 bp in length
* 52386 52485: gap of 100 bp
* 52486 53438: contig of 953 bp in length
* 53439 53538: gap of 100 bp
* 53539 54500: contig of 962 bp in length
* 54501 54600: gap of 100 bp
* 54601 55512: contig of 912 bp in length
* 55513 55612: gap of 100 bp
* 55613 56579: contig of 967 bp in length
* 56580 56679: gap of 100 bp
* 56680 57644: contig of 965 bp in length
* 57645 57744: gap of 100 bp
* 57745 58703: contig of 959 bp in length
* 58704 58803: gap of 100 bp
* 58804 59756: contig of 953 bp in length
* 59757 59856: gap of 100 bp
* 59857 60781: contig of 925 bp in length
* 60782 60881: gap of 100 bp
* 60882 61804: contig of 923 bp in length
* 61805 61904: gap of 100 bp

Alignment Scores:

Pred. No.: 7.32e-121 Length: 81583
Score: 1172.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-10-063-732-120 (1-225) x AC131927 (1-81583)

QY 1 MetAlaThrHisAlaLeuGluLeuAlaGlyLeuPheLeuGlyValGlyMetValGly 20
Db 36440 ATGGCAACCCATGCTTAGAAATCGCTGGGCTGTTCTTGGTGGTGGTGGTGGG 36381
QY 21 ThrValAlaValThrValMetProGlnTrpArgValSerAlaPheIleGluAsnAsnIle 40
Db 36380 ACAGTGGCTGTCACTGTCAATGCCCTCAGTGAGAGTGTGGCCTTCATTGAAACAAACATC 36321
QY 41 ValValPheGluAsnPheTrpGluGlyLeuTrpMetAsnCysValArgGlnAlaAsnIle 60
Db 36320 GTGGTTTTTGAAACCTTCTGGGAAGGACTGTGGATGAATTGCGTGAGGCGAGGTAAACATC 36261
QY 61 ArgMetGlnCysLysIleLeuTyrAspSerLeuLeuAlaLeuSerProAspLeuGlnAlaAla 80
Db 36260 AGGATGCAGTGCAAAATCTATGATTCCTGTGGCTCTTCTCCGGACCTACAGGCAGCC 36201
QY 81 ArgGlyLeuMetCysAlaAlaSerValMetSerPheLeuAlaPheMetMetAlaIleLeu 100
Db 36200 AGAGGACTGATGTGTGCTGCTTCCGTGATGTCCTTCTTGGCTTTCATGATGGCCATCCTT 36141
QY 101 GlyMetLysCysThrArgCysThrGlyAspAsnGluLysValLysAlaHisIleLeuLeu 120
Db 36140 GGCATGAAATGCACAGGTGCACGGGGGACAATGAGAAGGTGAAGGCTCACATTCTGCTG 36081
QY 121 ThrAlaGlyIleIlePheIleIleThrGlyMetValValLeuIleProValSerTrpVal 140
Db 36080 ACGGCTGGAATCATCTTTCATCATCACGGGCGATGGTGGTGTCTCATCCCTGTGAGCTGGGT 36021
QY 141 AlaAsnAlaIleIleArgAspPheTyrAsnSerIleValAsnValAlaGlnLysArgGlu 160
Db 36020 GCCAATGCCATCATCAGAGATTTCATAACTCAATAGTGAATGTTGCCCAAAACGTGAG 35961
QY 161 LeuGlyGluAlaLeuTyrLeuGlyTrpThrAlaLeuValLeuIleValGlyGlyAla 180
Db 35960 CTTGGAGAAGCTCTTACTTAGGATGACCAACGGCACTGGTGTGCTGATTGTTGGAGGAGCT 35901
QY 181 LeuPheCysCysValPheCysCysAsnGluLysSerSerTyrArgTyrSerIlePro 200

Db 35900 CTGTTCTGCTGCGTCTTTTGTGTCACGAAAGAGCAGTAGCTACAGATACCTCGATACCT 35841

QY 201 SerHisArgThrThrGlnLysSerTyrHisThrGlyLysLysSerProSerValTyrSer 220

Db 35840 TCCCATGCGCACACCCAAAAAGTTATCACACCGGAAAGAGTACCCGAGCGTCTACTCC 35781

QY 221 ArgSerGlnTyrVal 225

Db 35780 AGAAGTCAGTATGTG 35766

RESULT 9

AP000884/c

LOCUS

DEFINITION

AP000884

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. 137955

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q22.1"

/clone="B463J19"

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-10-063-732-120 (1-225) x AP000884 (1-137955)

QY 1 MetAlaThrHisAlaLeuGluTleAlaGlyLeuPheLeuGlyValGlyMetValGly 20

Db 58261 ATGGCAACCCATGCCTTAGAAATCGCTGGGCTGTTCTTGGTGGTGTGGAATGGTGGGC 58202

QY 21 ThrValAlaValThrValMetProGlnTrpArgValSerAlaPheIleGluAsnIle 40

Db 58201 ACAGTGGCTGTCACTGTCTCATGCCTCAGTGGAGAGTGTGGCCCTTCATGTGAAACACATC 58142

QY 41 ValValPheGluAsnPheTrpGluGlyLeuTrpMetAsnCysValArgGlnAlaAsnIle 60

Db 58141 GTGGTTTTTGAAGAACTTCTGGGAAGGACTGTGGATGAATTGCGTGAGCAGGCTTAATC 58082

QY 61 ArgMetGlnCysLysIleTyrAspSerLeuLeuAlaLeuSerProAspLeuGlnAlaAla 80

Db 58081 AGGATGCAGTGCAGAAATCTATGATTCCTGCTGGCTCTTTCTCCGAGCTACAGGCAGCC 58022

QY 81 ArgGlyLeuMetCysAlaAlaSerValMetSerPheLeuAlaPheMetMetAlaIleLeu 100

Db 58021 AGAGGACTGATGTGTGCTGCTCCGTGATGTCCTTCTTGGCTTTCATGATG3CCATCCTT 57962

QY 101 GlyMetLysCysThrArgCysThrGlyAspAsnGluLysValLysAlaHisIleLeuLeu 120

Db 57961 GGCATGAATGCACACGAGGTGCACGGGGGACAAATGAGAAGGTGAAGGCTCACATTCTGCTG 57902

QY 121 ThrAlaGlyIleIlePheIleIleThrGlyMetValValLeuIleProValSerTrpVal 140

Db 57901 ACGGCTGGAATCATCTTTCATCATCACGGGCATGGTGTGCTCATCCCTGTGAGCTGGTT 57842

QY 141 AlaAsnAlaIleIleArgAspPheTyrAsnSerIleValAsnValAlaGlnLysArgGlu 160

Db 57841 GCCAATGCCATCATCAGAGATTTCTATAACTCAATAGTGAATGTTGCCCAAAAACGTGAG 57782

QY 161 LeuGlyGluAlaLeuTyrLeuGlyTrpThrThrAlaLeuValLeuIleValGlyAla 180

Db 57781 CTTGGAGAAGCTCTCTACTTAGGATGGACCACGGCACTGGTGTGCTGATTGTTGGAGGAGCT 57722

QY 181 LeuPheCysCysValPheCysCysAsnGluLysSerSerSerTyrArgTyrSerIlePro 200

Db 57721 CTGTTCTGCTCGGTTTTTTGTTGCAACGAAAGAGGAGTAGCTACAGATACTCGATACCT 57662

QY 201 SerHisArgThrThrGlnLysSerTyrHisThrGlyLysLysSerProSerValTyrSer 220

Db 57661 TCCCATCGCACACCCAAAAAGTTATCACACCGGAAAGAGTCAACCGAGCGTCTACTCC 57602

QY 221 ArgSerGlnTyrVal 225

Db 57601 AGAAGTCAGTATGTG 57587

RESULT 10

AP001707/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1

Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Rosenthal,A., Kudoh,J., Shibuya,K., Schudy,A., Zimmermann,W., Rosenthal,A., Sasaki,K., Nagamine,K., Mitsuyama,S., Asakawa,S., Shintani,A., Sasaki,K., Shimizu,N., Nordsiek,G., Antonarakis,S.E., Minoshima,S., Scharfe,M., Schoen,O., Desario,A., Hornischer,K., Barandt,P., Scharfe,M., Ramser,J., Beck,A., Klages,S., Reichelt,J., Kauer,G., Bloecker,H., Dagand,E., Wehrmeyer,S., Borzym,K., Hennig,S., Rieselmann,L., Dagand,E., Lehrach,H., Reinhardt,R. and Yaspo,M.L.

The DNA sequence of human chromosome 21

Nature 405 (6784), 311-319 (2000)

20289799

10830953

2 (bases 1 to 340000)

Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Rosenthal,A., Kudoh,J., Shibuya,K., Schudy,A., Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Sasaki,K., Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S., Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G., Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S., Hennig,S., Rieselmann,L., Dagand,E., Wehrmaeyer,S., Borzym,K.,

Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Vasp0,M.L.
 Direct Submission
 Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)
 On Jan 16, 2002 this sequence version replaced gi:7717317.
 The Chromosome 21 mapping and sequencing consortium consisting of * RIKEN Genomic Sciences Center, Human Genome Research Group, * Sagami-hara 228-8555, Japan,
 * e.mail: hattori@gsc.riken.go.jp
 * URL: http://hgp.gsc.riken.go.jp/

and
 * Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany,
 * e.mail: gscj-submit@genome.imb-jena.de
 * URL: http://genome.imb-jena.de/

and
 * Keio University School of Medicine, Molecular Biology, * Tokyo 160-8582, Japan,
 * e.mail: nshimizu@dm-b-med.keio.ac.jp
 * URL: http://www.dmb.med.keio.ac.jp/
 and
 * GBF, Dept. of Genome Analysis,
 * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail: info.genome@gbf.de
 * URL: http://genome.gbf.de/

and
 * Max-Planck Institute for Molecular Genetics,
 * Ihnestrassse 73, D-14195 Berlin, Germany,
 * e.mail: info-chr21@molgen.mpg.de
 * URL: http://chr21.rz-berlin.mpg.de/
 AL163252: Submitted (10-Apr-2000).

FEATURES

source

1. 340000
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="21"
 <1..148
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.1"
 /clone="R355D16, 5' partial"
 /clone_lib="RPCI-11 BAC library"
 /note="Accession No. AP000236"

source

<1..48622
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.1"
 /clone="R347N6, 5' partial"
 /clone_lib="RPCI-11 BAC library"
 /note="Accession No. AP000390"

source

<1..57177
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.1"
 /clone="P681K20, 5' partial"
 /clone_lib="RPCI1,3-5 PAC library"
 /note="Accession No. AF191069"
 26572..135558
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

source

/chromosome="21"
 /map="21q22.1"
 /clone="P131P23"
 /clone_lib="RPCI1,3-5 PAC library"
 /note="Accession No. AF190641"
 48469..158332
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.1"
 /clone="P209K9"
 /clone_lib="RPCI1,3-5 PAC library"
 /note="Accession No. AF096876"
 156690..294644
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.1"
 /clone="R463J19"
 /clone_lib="RPCI-11 BAC library"
 /note="Accession No. AP000884"
 292534..>340000
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.1"
 /clone="R680H4, 3' partial"
 /clone_lib="RPCI-11 BAC library"
 /note="Accession No. AP000885"
 1610..1766
 /note="MER50"
 /rpt_family="LTR/MER4-group"
 /rpt_type=DISPERSED
 1857..2051
 /note="AluSx"
 /rpt_family="SINE/Alu"
 /rpt_type=DISPERSED
 2054..2392
 /note="MER50"
 /rpt_family="LTR/MER4-group"
 /rpt_type=DISPERSED
 3327..3518
 /note="MER102"
 /rpt_family="DNA/MER1_type"
 /rpt_type=DISPERSED
 complement(3582..3694)
 /note="L2"
 /rpt_family="LINE/L2"
 /rpt_type=DISPERSED
 4580..4614
 /note="(TTC)n"
 /rpt_family="Simple_repeat"
 /rpt_type=TANDEM
 5743..6052
 /note="AluSq"
 /rpt_family="SINE/Alu"
 /rpt_type=DISPERSED
 complement(6134..6214)
 /note="MIR"
 /rpt_family="SINE/MIR"
 /rpt_type=DISPERSED
 complement(6299..6651)
 /note="MLT1A1"
 /rpt_family="LTR/MaLR"
 /rpt_type=DISPERSED
 7531..7839
 /note="AluSc"
 /rpt_family="SINE/Alu"
 /rpt_type=DISPERSED
 9145..9173

repeat_region /note="(A)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(10285..10467)
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
10468..10753
/note="AluJb"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
10757..10785
repeat_region /note="(TAAAA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM

Alignment Scores:		4.44e-120	Length:	340000
Pred. No.:		1172.00	Matches:	225
Score:		100.00%	Conservative:	0
Percent Similarity:		100.00%	Mismatches:	0
Best local Similarity:		100.00%	Indels:	0
Query Match:		100.00%	Gaps:	0
DB:		9		
US-10-063-732-120 (1-225) x AP001707 (1-340000)				
QY	1	MetAlaThrHisAlaLeuGluIleAlaGlyLeuPheLeuGlyValGlyMetValGly	20	
Db	214950	ATGGAACCCATGCCTTAGAAATCGCTGGGCTGTTCTTGGTGGTGGTGGTGGC	214891	
QY	21	ThrValAlaValThrValMetProGlnTrpArgValSerAlaPheIleGluAsnIle	40	
Db	214890	ACAGTGGCTGTCACTGTCACTGCCTCAGTGGAGAGTGTGGCCTTCATTGAAAACAACATC	214831	
QY	41	ValValPheGluAsnPheTrpGluGlyLeuTrpMetAsnCysValArgGlnAlaAsnIle	60	
Db	214830	GTGGTTTGTGAAACTTCTGGGAAGGACTGTGGATGAATTGCGTGAGGCAGGCTAACATC	214771	
QY	61	ArgMetGlnCysLysIleTyrAspSerLeuLeuAlaLeuSerProAspLeuGlnAlaAla	80	
Db	214770	AGGATGCAGTGCAAAATCTATGATTCCCTGTGGCTCTTCTCCGGACCTACAGGCAGCC	214711	
QY	81	ArgGlyLeuMetCysAlaAlaSerValMetSerPheLeuAlaPheMetMetAlaIleLeu	100	
Db	214710	AGAGGACTGATGTGTGCTGCTCCGTGATGTCCTTCTTGGCTTCATGATGGCCATCCTT	214651	
QY	101	GlyMetLysCysThrArgCysThrGlyAspAsnGluLysValLysAlaHisIleLeuLeu	120	
Db	214650	GGCATGAATGCACCAAGGTGCACGGGGGACATGAGAAGGTGAAGGCTCACATTCGTCTG	214591	
QY	121	ThrAlaGlyIleIlePheIleIleThrGlyMetValValLeuIleProValSerTrpVal	140	
Db	214590	ACGGCTGGAATCATCTTCATCATCACGGGCATGGTGGTGTCTCATCCCTGTGAGCTGGGTT	214531	
QY	141	AlaAsnAlaIleIleArgAspPheTyrAsnSerIleValAsnValAlaGlnLysArgGlu	160	
Db	214530	GCCAATGCCATCATCAGAGATTCTATATACTCAATAGTAGTGAATGTTGCCCAAAAACGTGAG	214471	
QY	161	LeuGlyGluAlaLeuTyrLeuGlyTrpThrAlaLeuValLeuIleValGlyGlyAla	180	
Db	214470	CTTGGAGAAGCTCTCTACTTAGGATGGACCACGGCACTGGTGTCTGATTGTTGGAGGAGCT	214411	
QY	181	LeuPheCysCysValPheCysAsnGluLysSerSerTyrArgTyrSerIlePro	200	
Db	214410	CTGTTCTCTGTCGGTTTTTTTGTGCAACGAAAGACAGTAGCTACAGATACTCGATACT	214351	
QY	201	SerHisArgThrThrGlnLysSerTyrHisThrGlyLysLysSerProSerValTyrSer	220	
Db	214350	TCCCATCGACACACCCAAAAAGTTATCACACCGGAAAGAGTCAACCGGCGTCTACTCC	214291	
QY	221	ArgSerGlnTyrVal	225	
Db	214290	AGAAGTCAGTATGTG	214276	

RESULT 11
BS000178/c
LOCUS
DEFINITION
Pan troglodytes chromosome 22 clone:RP43-042C06, map 22, complete
sequences.
BS000178
BS000178.1 GI:37537445
HTG.
Pan troglodytes (chimpanzee)
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1
The Chimpanzee Chromosome 22 Sequencing Consortium.
DNA sequence of chimpanzee chromosome 22 and its evolutionary
implications
Unpublished
2 (bases 1 to 176580)
Wang,S., Cai,Z., Wang,B., Zheng,H., Zhang,Y., Zhang,X., Zhu,G.,
Lu,G., Fu,G. and Chen,Z.
Direct Submission
Submitted (26-MAY-2003) Shengyue Wang, Chinese National Human
Genome Center at Shanghai, Genomic Sequencing; No.250 BiBo Road,
Zhang Jiang HI-TECH Park, Shanghai 201203, CHINA
(E-mail:wangsy@chgc.sh.cn, URL:http://www.chgc.sh.cn,
Tel:86-21-50801919, Fax:86-21-50801922)
The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
*Chinese National Human Genome Center at Shanghai, Shanghai, China;
*GBF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
of Molecular Biotechnology, Jena, Germany; *KRIBB Genome Research
Center, Daejeon, Korea;
*Max-Planck-Institute for Molecular Genetics, Berlin, Germany;
*National Institute of Genetics, Mishima, Japan;
*National Yang Ming University Genome Research Center, Taipei,
Taiwan;
*RIKEN Genomic Sciences Center, Yokohama, Japan.
----- Genome Center
Center: Chinese National Human Genome Center at Shanghai Center
code: CHGCS
Web site: http://chgc.sh.cn
Contact: wangsy@chgc.sh.cn
----- Project Information
Center project name:The Chimpanzee Chromosome 22 Sequencing Project
Center clone name: RP43-042C06
----- Summary Statistics
Sequencing vector: pUC18,100% of reads
Chemistry: Dye-terminator Big Dye and ET; 100% of reads Assembly
Program: Phrap; version 0.990329
Consensus quality: 176318 bases at least Q40
Consensus quality: 176550 bases at least Q30
Consensus quality: 176577 bases at least Q20
Quality coverage: 9.0X

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30);
an attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by at one
plasmid
subclone or more than one M13 subclone;
and the assembly was confirmed by restriction digest.

Source information:
The RP43 chimpanzee BAC library was prepared from DNA isolated from
cultured cells established from the blood of a single male
chimpanzee.
Clones may be obtained from Asao Fujiyama and co-workers
(http://www.gsc.riken.go.jp).
VECTOR: pBACe3.6
Sequence Quality Assessment:
This entry has been annotated with sequence

estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

FEATURES
source
Neighboring clones: PTB-118H03(left) and RP43-082J09(right).
Location/Qualifiers
1. 176580
/organism="pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="22"
/clone="RP43-042C06"
/clone_lib="RP43 chimpanzee BAC"

ORIGIN

Alignment Scores:
Pred. No.: 7.17e-120 Length: 176580
Score: 1167.00 Matches: 224
Percent Similarity: 99.56% Conservative: 0
Best Local Similarity: 99.56% Mismatches: 1
Query Match: 99.57% Indels: 0
DB: 9 Gaps: 0
US-10-063-732-120 (1-225) x BS000178 (1-176580)
QY 1 MetAlaThrHisAlaLeuGluIleAlaGlyLeuPheLeuGlyGlyValGlyMetValGly 20
Db 31108 ATGGCAACCCAGCCTTAGAAATCGCTGGCTGTTTCTTGGTGTGGAATGGTGGC 31049
QY 21 ThrValAlaValThrValMetProGlnTrpArgValSerAlaPheIleGluAsnIle 40
Db 31048 ACAGTGGCTGTCACTGTCAATGCCCTCAGTGGAGAGTGTGGCCTTCATTGAAACAACATC 30989
QY 41 ValValPheGluAsnPheTrpGluGlyLeuTrpMetAsnCysValArgGlnAlaAsnIle 60
Db 30988 GTGGTTTTTGAAACTTCTGGGAAGGACTGTGGATGAATTGCTGAGCAGGCTAACATC 30929
QY 61 ArgMetGlnCysLysIleTyrAspSerLeuAlaLeuSerProAspLeuGlnAlaAla 80
Db 30928 AGGATGCAGTGCAGAAATCTATGATTCCTCGTGCTGCTTCTCCGGACCTACAGGCAGCC 30869
QY 81 ArgGlyLeuMetCysAlaAlaSerValMetSerPheLeuAlaPheMetAlaIleLeu 100
Db 30868 AGAGGACTGATGTGTGCTGCTTCGTGATGTCCTTCTTGGCTTTTCATGATGGCCATCCTT 30809
QY 101 GlyMetLysCysThrArgCysThrGlyAspAsnGluLysValLysAlaHisIleLeuLeu 120
Db 30808 GGCATGAAATGACCAAGGTGCACGGGGGACAATGAGAAGGTGAAGGCTCACATTCTGCTG 30749
QY 121 ThrAlaGlyIleIlePheIleIleThrGlyMetValValLeuIleProValSerTrpVal 140
Db 30748 ACGGCTGGAATCATCTTCATCATCAGGGCATGGTGTGCTCATCCCTGTGAGCTGGGT 30689
QY 141 AlaAsnAlaIleIleArgAspPheTyrAsnSerIleValAsnValAlaGlnLysArgGlu 160
Db 30688 GCCAATGCCATCATCAGAGATTCTATACCCAAATAGTGAATGTTGCCAAACGTGAG 30629
QY 161 LeuGlyGluAlaLeuTyrLeuGlyTrpThrThrAlaLeuValLeuIleValGlyGlyAla 180
Db 30628 CTGGAGAGAGCTCTACTTAGGATGGACCACCGGCACTGGTGTGATGTTGGAGGAGCT 30569
QY 181 LeuPheCysCysValPheCysCysAsnGluLysSerSerSerTyrArgTyrSerIlePro 200
Db 30568 CTGTTCTGCTGCGTTTTTTTGTTCACGAAAGACAGTAGCTACAGATACCTCGATACCT 30509
QY 201 SerHisArgThrThrGlnLysSerTyrHisThrGlyLysSerProSerValTyrSer 220
Db 30508 TCCCATCGCACACCCCAAAAAGTTATCACACCCGGAAGAAGTCACCGAGCGTCTACTCC 30449
QY 221 ArgSerGlnTyrVal 225
Db 30448 AGAAGTCAGTATGTG 30434

RESULT 12
BS000177/c

LOCUS
DEFINITION

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

BS000177 267172 bp DNA linear PRI 07-OCT-2003
Pan troglodytes chromosome 22 clone:PTB-118H03, map 22, complete
sequences.
BS000177
BS000177.1 GI:37537444
HTG.
Pan troglodytes (chimpanzee)
Pan troglodytes
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1 The Chimpanzee Chromosome 22 Sequencing Consortium.
DNA sequence of chimpanzee chromosome 22 and its evolutionary
implications
Unpublished
2 (bases 1 to 267172)
Wang,S., Cai,Z., Wang,B., Zheng,H., Zhang,Y., Zhang,X., Zhu,G.,
Lu,G., Fu,G. and Chen,Z.
Direct Submission
Submitted (26-MAY-2003) Shengyue Wang, Chinese National Human
Genome Center at Shanghai, Genomic Sequencing; No.250 BiBo Road,
Zhang Jiang HI-TECH Park, Shanghai 201203, CHINA
(E-mail:wangsy@chgc.sh.cn, URL:http://www.chgc.sh.cn,
Tel:86-21-50801919, Fax:86-21-50801922)
The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
*Chinese National Human Genome Center at Shanghai, Shanghai, China;
*GBF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
of Molecular Biotechnology, Jena, Germany; *KRIBB Genome Research
Center, Daejeon, Korea;
*Max-Planck-Institute for Molecular Genetics, Berlin, Germany;
*National Institute of Genetics, Mishima, Japan;
*National Yang Ming University Genome Research Center, Taipei,
Taiwan;
*RIKEN Genomic Sciences Center, Yokohama, Japan.
----- Genome Center
Center: Chinese National Human Genome Center at Shanghai Center
code: CHGCS
Web site: http://chgc.sh.cn
Contact: wangsy@chgc.sh.cn
----- Project Information
Center project name:The Chimpanzee Chromosome 22 Sequencing Project
Center clone name: PTB-118H03
----- Summary Statistics
Sequencing vector: pUC18,100% of reads
Chemistry: Dye-terminator Big Dye and ET; 100% of reads Assembly
program: Phrap; version 0.990329
Consensus quality: 266876 bases at least Q40
Consensus quality: 267150 bases at least Q30
Consensus quality: 267168 bases at least Q20
Quality coverage: 9.3x

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30);
an attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by at one
plasmid
subclone or more than one M13 subclone;
and the assembly was confirmed by restriction digest.

Source information:
The PTB1 chimpanzee BAC library was prepared from DNA isolated from
cultured cells established from the blood of a single male
chimpanzee.
Clones may be obtained from Asao Fujiyama and co-workers
(http://www.gsc.riken.go.jp).
VECTOR: pKS145
Sequence Quality Assessment:
This entry has been annotated with sequence

estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Neighboring clones: RP43-006021(left) and RP43-042C06(right).
Location/Qualifiers
1..267172
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="22"
/clone="PTB-118H03"
/clone_lib="PTB1 chimpanzee BAC"

FEATURES
source
1..267172
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="22"
/clone="PTB-118H03"
/clone_lib="PTB1 chimpanzee BAC"

ORIGIN
Alignment Scores:
Pred. No.: 121e-119 Length: 267172
Score: 1167.00 Matches: 224
Percent Similarity: 99.56% Conservative: 0
Best Local Similarity: 99.56% Mismatches: 1
Query Match: 99.57% Indels: 0
DB: 9 Gaps: 0

US-10-063-732-120 (1-225) x BS000177 (1-267172)

QY 1 MetAlaThrHisAlaLeuGluIleAlaGlyLeuPheLeuGlyGlyValGlyMetValGly 20
Db 258767 ATGGCAACCCATGCCCTAGAAATCGCTGGCTGTTCTTGGTGGTGTGGATGGTGGC 258708
QY 21 ThrValAlaValThrValMetProGlnTrpArgValSerAlaPheIleGluAsnAila 40
Db 258707 ACAGTGGCTGTCACTGTTCATGCCTCAGTGGAGAGTGTGGGCTTCATTGAAAACAACATC 258648
QY 41 ValValPheGluAsnPheTrpGluGlyLeuTrpMetAsnCysValArgGlnAlaAsnIle 60
Db 258647 GTGGTCTTTGAAAACCTCTGGGAAGGACCTGTGGATGAATTGGTGAGGACGAGCTAACATC 258588
QY 61 ArgMetGlnCysLysIleTyrAspSerLeuLeuAlaLeuSerProAspLeuGlnAlaAla 80
Db 258587 AGGATGCAGTGCAAAATCTATGATCCCTGGCTGCTCTTTCTCCGGACCTACAGGACGCC 258528
QY 81 ArgGlyLeuMetCysAlaAlaSerValMetSerPheLeuAlaPheMetMetAlaIleLeu 100
Db 258527 AGAGGACTGATGTGTGCTGCTTCCGTGATGCTCTTCTTGGCTTTCATGATGGCCATCCTT 258468
QY 101 GlyMetLysCysThrArgCysThrGlyAspAsnGluLysValLysAlaHisIleLeuLeu 120
Db 258467 GGCATGAAATGCACAGTGCACGGGGACAATGAGAAGGTGAAGGCTCACATTCTGCTG 258408
QY 121 ThrAlaGlyIleIlePheIleIleThrGlyMetValValLeuIleProValSerTrpVal 140
Db 258407 ACGGCTGGAATCATCTTTCATCATCACGGGCATGGTGGTCTCATCCCTGTGAGCTGGGTT 258348
QY 141 AlaAsnAlaIleIleArgAspPheTyrAsnSerIleValAsnValAlaGlnLysArgGlu 160
Db 258347 GCCATGCCATCATCAGAGATTCTATAACCAATAGTGAATGTGCCCCAAAACGTGAG 258288
QY 161 LeuGlyGluAlaLeuTyLeuGlyTrpThrThrAlaLeuValLeuIleValGlyGlyAla 180
Db 258287 CTTGGAGAAGCTCTCTACTTAGGATGGACCGGCACCTGGTGTGTGATTGTTGGAGAGCT 258228
QY 181 LeuPheCysCysValPheCysCysAsnGluLysSerSerSerTyrArgTyrSerIlePro 200
Db 258227 CTGTTCTGCTGCGTTTTTGTGCAACGAAAGAGCAGTAGCTACAGATACTCGATACCT 258168
QY 201 SerHisArgThrThrGlnLysSerTyrHisThrGlyLysLysSerProSerValTyrSer 220
Db 258167 TCCCATGCGACAACCCAAAAGTTATCACACCGGAAAGAGTACCCGAGCGTCTACTCC 258108
QY 221 ArgSerGlnTyrVal 225
Db 258107 AGAAGTCAGTATGTG 258093

RESULT 13
AX866683
LOCUS AX866683 837 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 1588 from Patent EP1074617.
ACCESSION AX866683
VERSION AX866683.1 GI:40021032
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
TITLE Primers for synthesising full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 1588 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
source
1..837
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 9.72e-120 Length: 837
Score: 1140.00 Matches: 221
Percent Similarity: 98.23% Conservative: 1
Best Local Similarity: 97.79% Mismatches: 3
Query Match: 97.27% Indels: 1
DB: 6 Gaps: 0

US-10-063-732-120 (1-225) x AX866683 (1-837)

QY 1 MetAlaThrHisAlaLeuGluIleAlaGlyLeuPheLeuGlyGlyValGlyMetValGly 20
Db 150 ATGGCAACCCATGCCCTAGAAATCGCTGGCTGTTCTTGGTGGTGTGGATGGTGGC 209
QY 21 ThrValAlaValThrValMetProGlnTrpArgValSerAlaPheIleGluAsnAila 40
Db 210 ACAGTGGCTGTCACTGTTCATGCCTCAGTGGAGAGTGTGGGCTTCATTGAAAACAACATC 269
QY 41 ValValPheGluAsnPheTrpGluGlyLeuTrpMetAsnCysValArgGlnAlaAsnIle 60
Db 270 GTGGTCTTTGAAAACCTCTGGGAAGGACCTGTGGATGAATTGGTGGGACGCTAACATC 329
QY 61 ArgMetGlnCysLysIleTyrAspSerLeuLeuAlaLeuSerProAspLeuGlnAlaAla 80
Db 330 AGGATGCAGTGCAAAATCTATGATCCCTGCTGCTCTTCTCCGGACCTACAGGACGCC 389
QY 81 ArgGlyLeuMetCysAlaAlaSerValMetSerPheLeuAlaPheMetMetAlaIleLeu 100
Db 390 AGAGGACTGATGTGTGCTGCTTCCGTGATGCCCTTCTTGGCTTTCATGATGGCCATCCTT 449
QY 101 GlyMetLysCysThrArgCysThrGlyAspAsnGluLysValLysAlaHisIleLeuLeu 120
Db 450 GGCATGAAATGCACAGTGCACGGGGACAATGAGAAGGTGAAGGCTCACATTCTGCTG 509
QY 121 ThrAlaGlyIleIlePheIleIleThrGlyMetValValLeuIleProValSerTrpVal 140
Db 510 ACGGCTGGAATCATCTTTCATCATCACGGGCATGGTGGTGTCTCATCCCTGTGAGCTGGGTT 569
QY 141 AlaAsnAlaIleIleArgAspPheTyrAsnSerIleValAsnValAlaGlnLysArgGlu 160
Db 570 GCCAATGCCATCATCAGAGATTCTATAACCAATAGTGAATGTTGCCAAAACGTGAG 629
QY 161 LeuGlyGluAlaLeuTyLeuGlyTrpThrThrAlaLeuValLeuIleValGlyGlyAla 180
Db 630 CTTGGAGAAGCTCTCTACTTAGGATGGACCGGCACGCTGGTGTGCTGATGTTGGAGAGCT 689
QY 181 LeuPheCysCysValPheCysCysAsnGluLysSerSerSerTyrArgTyrSerIlePro 200
Db 181 LeuPheCysCysValPheCysCysAsnGluLysSerSerSerTyrArgTyrSerIlePro 200

Db 690 CTGTTCTGCTGCGTTTCTTTGTTGCAACGAAAGAGCAGTAGCTACAGATACTCGATACCT 749

Qy 201 SerHisArgThrThrGlnLysSerTyrHisThrGlyLysLysSer-ProSerValTyrSe 220

Db 750 TCCCATCGCACACCAAAAGTTATCACACCGGAAAGATCACCCGACCGTNTACTT 809

Qy 220 rArgSerGlnTyrVal 225

Db 810 CAGAAAGTCAGTATGTG 825

RESULT 14

BD146745

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FT source

FT

FEATURES

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-10-063-732-120 (1-225) x BD146745 (1-837)

Qy 1 MetAlaThrHisAlaLeuGluLeuAlaGlyLeuPheLeuGlyGlyValGlyMetValGly 20

Db 150 ATGGCAACCCATGCGCTTAGAAATCGCTGGGCTGTTCTTGGTGGTGGTGGTGGG 209

Qy 21 ThrValAlaValThrValMetProGlnTyrArgValSerAlaPheIleGluAsnIle 40

Db 210 ACAGTGGCTGTCACTGTCTATGCGCTCAGTGGAGAGTGTGGCTTTCATTGAAACAACATC 269

Qy 41 ValValPheGluAsnPheTyrGluGlyLeuTyrMetAsnGlnValArgGlnAlaAsnIle 60

Db 270 GTGGTGTGTTGAAAACTTCTGGGAAGGACTGTGGATGAATTGCGTGAGGCAGGCTAACATC 329

Qy 61 ArgMetGlnCysLysIleTyrAspSerLeuLeuAlaLeuSerProAspLeuGlnAlaAla 80

Db 330 AGGATGCAGTGCAAAATCTATGATTCCCTGCTGGCTCTTTCTCCGACCTACAGGCAGCC 389

Qy 81 ArgGlyLeuMetCysAlaAlaSerValMetSerPheLeuAlaPheMetMetAlaIleLeu 100

Db 390 AGAGGACTGATGTGTGCTGCTTCCGTGATGCCCTTCTTGGCTTTTCATGATGGCCATCCTT 449

Qy 101 GlyMetLysCysThrArgCysThrGlyAspAsnGlnLysValLysAlaHisIleLeuLeu 120

Db 450 GGCATGAAATGCACCAAGGTGCACGGGGACAATGAGAAGGTGAAGGCTCACATTCTGCTG 509

Qy 121 ThrAlaGlyIleIlePheIleIleThrGlyMetValValLeuIleProValSerTrpVal 140

Db 510 ACGGCTGGAATCATCTTTCATCATCACGGGCATGGTGGTGTCTATCCCTGTGAGCTGGGT 569

Qy 141 AlaAsnAlaIleIleArgAspPheTyrAsnSerIleValAsnValAlaGlnLysArgGlu 160

Db 570 GCCAATGCCATCATCAGAGATTCTATAACCCCAATAGTAATGTTGCCCAAAACGTGAG 629

Qy 161 LeuGlyGluAlaLeuTyrLeuGlyTyrThrThrAlaLeuValLeuIleValGlyGlyAla 180

Db 630 CTGGAGAGCTCTCTACTTAGGATGGACCAACGGCACTGGTGTCTGATTCTGGAGGAGCT 689

Qy 181 LeuPheCysCysValPheCysCysAsnGluLysSerSerSerTyrArgTyrSerIlePro 200

Db 690 CTGTTCTGCTGCGTTTTTTGTTGCAACGAAAGAGCAGTAGCTACAGATACTCGATACCT 749

Qy 201 SerHisArgThrThrGlnLysSerTyrHisThrGlyLysLysSer-ProSerValTyrSe 220

Db 750 TCCCATCGCACACCAAAAGTTATCACACCGGAAAGATCACCCGAAAGTCAACCCGACCGTNTACTT 809

Qy 220 rArgSerGlnTyrVal 225

Db 810 CAGAAAGTCAGTATGTG 825

RESULT 15

AP001846

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AP001846

191923 bp DNA linear HTG 30-MAY-2000

Homo sapiens chromosome 18 clone RP11-796C24 map 18q22, WORKING

DRAFT SEQUENCE, 56 unordered pieces.

AP001846

AP001846.2 GI:8117509

HTG; HTGS_PHASE1; HTGS_DRAFT.

Homo sapiens (human)

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 191923)

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

Homo sapiens 191,923 genomic DNA of 18q22

Published Only in DataBase (2000)

2 (bases 1 to 191923)

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

Direct Submission

Submitted (18-APR-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp, Fax:81-42-778-9924)

On May 30, 2000 this sequence version replaced gi:7630218.

----- Genome Center

Center: RIKEN Genomic Sciences Center (GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraft18

Center clone name: RP11-796C24

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator BT-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 167257 bases at least Q40
Consensus quality: 178235 bases at least Q30
Consensus quality: 183393 bases at least Q20
Insert size: 186423; sum-of-contigs
Quality coverage: 4.34x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 56 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1	10920	contig of	10920	bp in length
11021	18635	contig of	7615	bp in length
18736	27618	contig of	8883	bp in length
27719	33370	contig of	5652	bp in length
33471	38522	contig of	5052	bp in length
38623	44935	contig of	6313	bp in length
45036	52903	contig of	7868	bp in length
53004	59933	contig of	6930	bp in length
60034	65715	contig of	5682	bp in length
65816	73615	contig of	7800	bp in length
73716	79180	contig of	5465	bp in length
79281	83364	contig of	4084	bp in length
83465	87054	contig of	3590	bp in length
87155	91199	contig of	4045	bp in length
91300	94966	contig of	3667	bp in length
95067	99407	contig of	4341	bp in length
99508	103203	contig of	3696	bp in length
103304	106888	contig of	3585	bp in length
106989	110187	contig of	3199	bp in length
110288	113798	contig of	3511	bp in length
113899	117663	contig of	3765	bp in length
117764	121135	contig of	3372	bp in length
121236	123224	contig of	1989	bp in length
123325	125604	contig of	2280	bp in length
125705	127990	contig of	2286	bp in length
128091	130683	contig of	2593	bp in length
130784	133371	contig of	2588	bp in length
133372	133471	contig of	2955	bp in length
133472	136426	contig of	2955	bp in length
136427	136526	contig of	2730	bp in length
136527	139256	contig of	2730	bp in length
139257	139356	contig of	2032	bp in length
139357	141388	contig of	2032	bp in length
141389	141489	contig of	2601	bp in length
141489	144089	contig of	2601	bp in length
144090	144189	contig of	2497	bp in length
144190	146686	contig of	2497	bp in length
146687	146786	contig of	3087	bp in length
146787	149873	contig of	3087	bp in length
149874	149973	contig of	3087	bp in length

Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently

* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1	10920	contig of	10920	bp in length
10921	11020	gap of	100	bp
11021	18635	contig of	7615	bp in length
18636	18735	gap of	100	bp
18736	27618	contig of	8883	bp in length
27619	27718	gap of	100	bp
27719	33370	contig of	5652	bp in length
33371	33470	gap of	100	bp
33471	38522	contig of	5052	bp in length
38523	38622	gap of	100	bp
38623	44935	contig of	6313	bp in length
44936	45035	gap of	100	bp
45036	52903	contig of	7868	bp in length
52904	53003	gap of	100	bp
53004	59933	contig of	6930	bp in length
59934	60033	gap of	100	bp
60034	65715	contig of	5682	bp in length
65716	73615	contig of	7800	bp in length
73616	73715	gap of	100	bp
73716	79180	contig of	5465	bp in length
79181	79280	gap of	100	bp
79281	83364	contig of	4084	bp in length
83365	83464	gap of	100	bp
83465	87054	contig of	3590	bp in length
87055	87154	gap of	100	bp
87155	91199	contig of	4045	bp in length
91200	91299	gap of	100	bp
91300	94966	contig of	3667	bp in length
94967	95066	gap of	100	bp
95067	99407	contig of	4341	bp in length
99408	99507	gap of	100	bp
99508	103203	contig of	3696	bp in length
103204	103303	gap of	100	bp
103304	106888	contig of	3585	bp in length
106889	106988	gap of	100	bp
106989	110187	contig of	3199	bp in length
110188	110287	gap of	100	bp
110288	113798	contig of	3511	bp in length
113799	113898	gap of	100	bp
113899	117663	contig of	3765	bp in length
117664	117763	gap of	100	bp
117764	121135	contig of	3372	bp in length
121136	121235	gap of	100	bp
121236	123224	contig of	1989	bp in length
123225	123324	gap of	100	bp
123325	125604	contig of	2280	bp in length
125605	125704	gap of	100	bp
125705	127990	contig of	2286	bp in length
127991	128090	gap of	100	bp
128091	130683	contig of	2593	bp in length
130684	130783	gap of	100	bp
130784	133371	contig of	2588	bp in length
133372	133471	gap of	100	bp
133472	136426	contig of	2955	bp in length
136427	136526	gap of	100	bp
136527	139256	contig of	2730	bp in length
139257	139356	gap of	100	bp
139357	141388	contig of	2032	bp in length
141389	141488	gap of	100	bp
141489	144089	contig of	2601	bp in length
144090	144189	gap of	100	bp
144190	146686	contig of	2497	bp in length
146687	146786	gap of	100	bp
146787	149873	contig of	3087	bp in length
149874	149973	gap of	100	bp

* 149974 151895: contig of 1922 bp in length
* 151896 151995: gap of 100 bp
* 151996 154840: contig of 2845 bp in length
* 154841 154940: gap of 100 bp
* 154941 157109: contig of 2169 bp in length
* 157110 157209: gap of 100 bp
* 157210 159650: contig of 2441 bp in length
* 159651 159750: gap of 100 bp
* 159751 161167: contig of 1417 bp in length
* 161168 161267: gap of 100 bp
* 161268 163443: contig of 2176 bp in length
* 163444 163543: gap of 100 bp
* 163544 165635: contig of 2092 bp in length
* 165636 165735: gap of 100 bp
* 165736 168432: contig of 2697 bp in length
* 168433 168532: gap of 100 bp
* 168533 171331: contig of 2799 bp in length
* 171332 171431: gap of 100 bp

Alignment Scores:

Pred. No.: 4.45e-116 Length: 191923
Score: 1134.00 Matches: 222
Percent Similarity: 98.67% Conservative: 0
Best Local Similarity: 98.67% Mismatches: 3
Query Match: 96.76% Indels: 1
DB: 2 Gaps: 0

US-10-063-732-120 (1-225) x AP001846 (1-191923)

Qy 1 MetAlaThrHisAlaLeuGluLeuAlaGlyLeuPheLeuGlyGlyValGlyMetValGly 20
Db 129410 ATGGCAACCCATGCCCTTAAA-ATCGCTGGGCTGTTTCTTGGTGGTGGTGGTGGGC 129468
Qy 21 ThrValAlaValThrValMetProGlnTrpArgValSerAlaPheIleGluAsnAsnIle 40
Db 129469 ACAGTGGCTGTCACTGTCTATGCTCAGTGGAGAGTGTGGCCTTCATTGAAACACATC 129528
Qy 41 ValValPheGluAsnPheTrpGluGlyLeuTrpMetAsnCysValArgGlnAlaAsnIle 60
Db 129529 GTGGTTTTTGAAAACTTCTGGGAAGGACTGTGGATGAATTGCGTGAGGAGGTAACATC 129588
Qy 61 ArgMetGlnCysLysIleTyrAspSerLeuLeuAlaLeuSerProAspLeuGlnAlaAla 80
Db 129589 AGGATGCAGTGCAGAAATCTATGATTCCTGTGGCTCTTTCTCCGGACCTACAGGCAGCC 129648
Qy 81 ArgGlyLeuMetCysAlaAlaSerValMetSerPheLeuAlaPheMetMetAlaIleLeu 100
Db 129649 AGAGGACTGATGTGTGCTTCCGTGATGTCTTGGCTTTTCATGATGGCCATCCTT 129708
Qy 101 GlyMetLysCysThrArgCysThrGlyAspAsnGluLysValLysAlaHisIleLeuLeu 120
Db 129709 GGCATGAATGACCAAGGTGCACGGGGGACAAATGAGAAGGTGAAGGCTCACATTCTGCTG 129768
Qy 121 ThrAlaGlyIleIlePheIleIleThrGlyMetValValLeuIleProValSerTrpVal 140
Db 129769 ACGGCTGGAATCATCTTCATCATCGCGGGCATGGTGGTGTCTCATCCCTGTGAGCTGGGTT 129828
Qy 141 AlaAsnAlaIleIleArgAspPheTyrAsnSerIleValAsnValAlaGlnLysArgGlu 160
Db 129829 GCCAATGCCATCATCAGAGATTCTATACCCCAATAGTGAATGTTGGCCCAAAACGTGAG 129888
Qy 161 LeuGlyGluAlaLeuTyrLeuGlyTrpThrThrAlaLeuValLeuIleValGlyGlyAla 180
Db 129889 CTTGGAGAAGCTCTCTACTTAGATGGACCCACGGGACTGGTGTGATGTTGGAGGAGCT 129948
Qy 181 LeuPheCysCysValPheCysCysAsnGluLysSerSerSerTyrArgTyrSerIlePro 200
Db 129949 CTGTTCTGTGCGGTTTTTTTGTGGCAACGAAAGAGCAGTAGCTACAGATACTCGATACCT 130008
Qy 201 SerHisArgThrThrGlnLysSerTyrHisThrGlyLysLysSerProSerValTyrSer 220
Db 130009 TCCCATCGCACAAACCCAAAAAGTTATCACCCGGAAGAGTACCGAGCGGTCTACTCC 130068

Qy 221 ArgSerGlnTyrVal 225
Db 130069 AGAAGTCAGTATGTG 130083
Search completed: September 3, 2004, 04:36:28
Job time : 2908 secs